

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 29, 2004, 10:21:25 ; Search time 2490 Seconds
(without alignments)
1235.263 Million cell updates/sec

Title: US-09-844-353A-54
Perfect score: 555
Sequence: 1 KKTITRRNAGNMSYALIT.....SSWWINPDAGMPPRRTR 103

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n_model -DEV=xlpl
-Q/cgn2_1/USPTO.spool_p/US09844353/runat_23072004 084914 28247/app_query.fasta_1.263
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosun62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=oct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTEWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09844353 @CGN_1_1_6425 @runat_23072004 084914 28247 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	478	86.1	594	12	BM292020	BM292020 EST574562
2	475	85.6	718	12	BM290603	BM290603 EST577137
3	472	85.0	491	12	BG226380	BG226380 Kq20R09.Y
4	469	84.5	590	12	BM291076	BM291076 EST577610
5	453	81.6	515	12	BM291800	BM291800 EST574342
6	440.5	79.4	593	13	BX673604	BX673604 BX673604
7	440.5	79.4	600	12	BM251446	BM251446 EOTL01000
8	440.5	79.4	628	10	BE881329	BE881329 601490517
9	440.5	79.4	998	11	BC019532	BC019532 Mus muscu
10	440.5	79.4	2530	11	AK047413	AK047413 Mus muscu
11	427.5	77.0	448	9	AV616467	AV616467 AV616467
12	427.5	77.0	459	10	AW204560	AW204560 UI-H-BL1-
13	427.5	77.0	492	12	BI181558	BI181558 UNL-P-FN-
14	427.5	77.0	505	9	AI804396	AI804396 tc62909.x
15	427.5	77.0	545	9	AI382221	AI382221 tc04605.x
16	427.5	77.0	553	9	AI698758	AI698758 wa81d11.x
17	427.5	77.0	556	10	AW298104	AW298104 UI-H-BW0-
18	427.5	77.0	593	9	AI766667	AI766667 w102806.x
19	427.5	77.0	681	10	BF057343	BF057343 7k19e10.x
20	427.5	77.0	716	13	BQ603828	BQ603828 MI-P-CPI-
21	427.5	77.0	722	9	AI439635	AI439635 tc91d08.x
22	427.5	77.0	741	10	BF058976	BF058976 7k36f09.x
23	427.5	77.0	753	9	AI808294	AI808294 wf54507.x
24	427.5	77.0	822	9	AI418347	AI418347 tg48b07.x
25	427.5	77.0	848	9	AI523901	AI523901 t997g10.x
26	427.5	77.0	870	12	BI182367	BI182367 UNL-P-FN-
27	427.5	77.0	883	13	BQ604124	BQ604124 MI-P-CPI-
28	427.5	77.0	894	9	AI472322	AI472322 t387c05.x
29	427.5	77.0	896	12	BI183684	BI183684 UNL-P-FN-
30	425.5	76.7	580	9	AI761654	AI761654 wf59a04.x
31	425.5	76.7	743	13	BX297137	BX297137 BX297137
32	423.5	76.3	648	10	BF115229	BF115229 7n80b07.x
33	422.5	76.1	613	9	AI221705	AI221705 qh05a04.x
34	421.5	75.9	608	10	AW573095	AW573095 hf31h12.x
35	421.5	75.9	612	10	BF057618	BF057618 7k46h02.x
36	421.5	75.9	621	10	AW517242	AW517242 xp92g08.x
37	421.5	75.9	686	10	BF057595	BF057595 7k46b06.x
38	421.5	75.9	757	9	AI813640	AI813640 wj65a09.x
39	421.5	75.9	793	12	BI183274	BI183274 UNL-P-FN-
40	421.5	75.9	813	12	BI184990	BI184990 UNL-P-FN-
41	421	75.9	397	14	CB772145	CB772145 AMGNNUC.T
42	419.5	75.6	422	9	AA769305	AA769305 nz39e07.s
43	419.5	75.6	685	14	CA316065	CA316065 UI-M-FW0-
44	418.5	75.4	577	9	AW028783	AW028783 ww35b10.x
45	418.5	75.4	677	12	BI185172	BI185172 UNL-P-FN-

ALIGNMENTS

RESULT 1
BM292020
LOCUS BM292020
DEFINITION AvSG Amblyomma variegatum cDNA clone AVABF26 5' end, mRNA
ACCESSION BM292020
VERSION BM292020
KEYWORDS EST.
SOURCE Amblyomma variegatum
ORGANISM Amblyomma variegatum
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Parasitiformes; Ixodida; Ixodidae; Amblyomma.
REFERENCE 1 (bases 1 to 594)

AUTHORS Nene,V., Lee,D., Quackenbush,J., Skilton,R., Mwaura,S., Gardner,M.J. and Bishop,R.
 TITLE AVGI, an index of genes transcribed in the salivary glands of the ixodid tick Amblyomma variegatum
 JOURNAL Int. J. Parasitol. 32 (12), 1447-1456 (2002)
 MEDLINE 22281296
 PUBMED 12392910
 COMMENT Contact: Vish Nene
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-610-5968
 Fax: 301-838-0208
 Email: nene@tigr.org
 Seq primer: M13 reverse.
 Location/Qualifiers
 1. .594
 /organism="Amblyomma variegatum"
 /mol_type="mRNA"
 /db_xref="taxon:34610"
 /clone="AVABF26"
 /tissue type="Salivary glands"
 /dev stage="Adult"
 /lab host="E.coli strain DH10B-Tona"
 /clone lib="AvSG"
 /note="Vector: pCMV-SPORT6.1; Salivary glands were dissected on day five after initiation of feeding. Total RNA was prepared using acid guanidium thiocyanate-phenol-chloroform extraction. The cDNA library was custom prepared by Invitrogen Corporation. Briefly, a first strand cDNA was primed using oligo(dT) containing a NotI site. Size fractionated double stranded cDNA was ligated to EcoRV-NotI cleaved vector and electroporated into E.coli."
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.5e-47 Length: 594
 Score: 478.00 Matches: 84
 Percent Similarity: 94.06% Conservative: 11
 Best Local Similarity: 83.17% Mismatches: 6
 Query Match: 86.13% Indels: 0
 DB: 12 Gaps: 0
 US-09-844-353A-54 (1-103) x BM292020 (1-594)
 Qy 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
 Db 259 AAAAAGAACTCGTCGCTGCGCAACGCTGGGGCAACATGCTCTAGCGGACCTGATTACG 318
 Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrIleuAlaGlnValTyrGluTrpMet 40
 Db 319 CAAGCCATCCAAAGCGCGCCGAGAAACGCTCAGCTGTCTACAGATCTACAGTGGATG 378
 Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTyrLys 60
 Db 379 GTGAGAACGTCCTCTATTTCAGGACAAAGGGCGACACACAGCTCCGCGCTGGAG 438
 Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
 Db 439 AACTCTATCCGGCACAACTGTGCTGTCACAGCAGATTCATGCGAGTCCAGAACCAAGGC 498
 Qy 81 AlaGlyLysSerSerTyrTrpValIleAsnProAspAlaLysProGlyMetAsnProArg 100
 Db 499 GCCGCAAGAGCTCGTGGTGGATGCTGAACCCGAGCGCAAGCCCGGCGAGCGCGAGG 558
 Qy 101 Arg 101
 Db 559 CGC 561
 RESULT 2
 BM290603
 LOCUS 718 bp mRNA linear EST 01-JUL-2002

DEFINITION EST577137 AvSG Amblyomma variegatum cDNA clone AVAAO69 5' end, mRNA sequence.
 ACCESSION BM290603
 VERSION BM290603.1 GI:21640570
 KEYWORDS EST.
 SOURCE Amblyomma variegatum
 ORGANISM Amblyomma variegatum
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari; Parasitiformes; Ixodida; Ixodidae; Amblyomma.
 REFERENCE 1 (bases 1 to 718)
 AUTHORS Nene,V., Lee,D., Quackenbush,J., Skilton,R., Mwaura,S., Gardner,M.J. and Bishop,R.
 TITLE AVGI, an index of genes transcribed in the salivary glands of the ixodid tick Amblyomma variegatum
 JOURNAL Int. J. Parasitol. 32 (12), 1447-1456 (2002)
 MEDLINE 22281296
 PUBMED 12392910
 COMMENT Contact: Vish Nene
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-610-5968
 Fax: 301-838-0208
 Email: nene@tigr.org
 Seq primer: M13 reverse.
 Location/Qualifiers
 1. .718
 /organism="Amblyomma variegatum"
 /mol_type="mRNA"
 /db_xref="taxon:34610"
 /clone="AVAAO69"
 /tissue type="Salivary glands"
 /dev stage="Adult"
 /lab host="E.coli strain DH10B-Tona"
 /clone lib="AvSG"
 /note="Vector: pCMV-SPORT6.1; Salivary glands were dissected on day five after initiation of feeding. Total RNA was prepared using acid guanidium thiocyanate-phenol-chloroform extraction. The cDNA library was custom prepared by Invitrogen Corporation. Briefly, a first strand cDNA was primed using oligo(dT) containing a NotI site. Size fractionated double stranded cDNA was ligated to EcoRV-NotI cleaved vector and electroporated into E.coli."
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.66e-47 Length: 718
 Score: 475.00 Matches: 84
 Percent Similarity: 93.07% Conservative: 10
 Best Local Similarity: 83.17% Mismatches: 7
 Query Match: 85.59% Indels: 0
 DB: 12 Gaps: 0
 US-09-844-353A-54 (1-103) x BM290603 (1-718)
 Qy 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
 Db 307 AAAAAGAACTCGTCGCTGCGCAACGCTGGGGCAACATGCTCTAGCGGACCTGATTACG 366
 Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrIleuAlaGlnValTyrGluTrpMet 40
 Db 367 CAAGCCATCCAAAGCGCGCCGAGAAACGCTCAGCTGTCCACAGATCTACAGTGGATG 426
 Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTyrLys 60
 Db 427 GTGAGAACGTCGCTCTATTTCAGGACAAAGGGCGACACACAGCTCCGCGCTGGAG 486
 Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
 Db 487 AACTCTATCCGGCACAACTGTGCTGTCACAGCAGATTCATGCGAGTCCAGAACCAAGGC 546
 Qy 81 AlaGlyLysSerSerTyrTrpValIleAsnProAspAlaLysProGlyMetAsnProArg 100

Db 547 GCCGCAAGAGCTGTTGGATGTGAACCGGACGCGCAAGCGCGGCGGAGG 605
 Qy 101 Arg 101
 Db 607 CGC 609

RESULT 3
 BG226380
 LOCUS
 DEFINITION
 TR_016849.v1.TN95TM-SSR Strongyloides stercoralis cDNA 5' similar to
 TF_016849.O16849 FORK HEAD-RELATED TRANSCRIPTION FACTOR DAF-16A2.
 [1], mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BG226380 491 bp mRNA linear EST 09-MAY-2001
 BG226380
 BG226380.1 GI:12713935
 EST.
 Strongyloides stercoralis
 Strongyloides stercoralis
 Strongyloides stercoralis
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Panagrolaimidae; Strongyloidea; Strongyloides.
 1 (bases 1 to 491)
 McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
 Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
 Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
 Tagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
 Underwood,K., Seftoe,M., Allen,M., Person,B., Swaller,T.,
 Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
 McCann,R., Waterston,R. and Wilson,R.
 The Washington Univ. Nematode EST Project, 1999
 Unpublished (1999)
 Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

TITLE
 JOURNAL
 COMMENT

The library was constructed by Dr. Thomas Nutman and colleagues of
 NIAID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
 University Genome Sequencing Center St. Louis.
 High quality sequence stop: 338.
 Location/Qualifiers
 1..491
 /organism="Strongyloides stercoralis"
 /mol_type="mRNA"
 /strain="Rhabditiform larvae obtained from gerbils"
 /db_xref="taxon:6248"
 /lab_host="XL-1 Blue MRF" (Stratagene)
 /clone_lib="TN95TM-SSR"
 /note="Vector: lambda Uni-ZAP XR (Stratagene); Site 1:
 EcoRI; Site 2: XhoI; mRNA was purified from 2 x 10E3
 rhabditiform larvae which had been isolated from gerbils
 experimentally infected with larvae originally isolated
 from experimentally infected dogs. cDNA was constructed
 and, using adaptors, was cloned unidirectionally into the
 vector from the EcoRI site to the XhoI site. The library
 has an unamplified titer of 1 x 10E5 pfu/ml and an
 amplified, undiluted titer of 9 x 10E11 pfu/ml. The
 average insert size of the unamplified library is 675 bp
 (range, 100-1700)."

ORIGIN

Alignment Scores:
 Pred. No.: 5,96e-47 Length: 491
 Score: 472.00 Matches: 87
 Percent Similarity: 92.23% Conservative: 8
 Best Local Similarity: 84.47% Mismatches: 8
 Query Match: 85.05% Indels: 0
 DB: 12 Gaps: 0

US-09-844-353A-54 (1-103) x BG226380 (1-491)

Oy 1 LysElyThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuThr 20

Pred. No.: 1.84e-46 Length: 590
 Score: 469.00 Matches: 83
 Percent Similarity: 93.07% Conservative: 11
 Best Local Similarity: 82.18% Mismatches: 7
 Query Match: 84.50% Indels: 0
 DB: 12 Gaps: 0

US-09-844-353A-54 (1-103) x BM291076 (1-590)

QY 1 LysLysThrThrArgArgAsnAlaTrpGlyAsnMetSertYrAlaGluLeuIleThr 20
 Db 255 AAAAGAACTCGTGGCTCGCAACGCTGGGGCAACATGCTCTACGGGACCTGATTACG 314
 QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
 Db 315 CAAGCCATCCAAAGCGCGCGAGAAACGCGCTCTACAGATCTACGAGTGGATG 374
 QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
 Db 375 GTGCAGAACGTGCGCTATTTCAGAGACAGAGGCGACAGACAGCTCGCGCGTGGAG 434
 QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
 Db 435 AACTCTATCCGGCAACACCTGCTGTCACAGCAGATTTCATGCGAGTCCAGAACGAAGC 494
 QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAspAlaLysProGlyMetAsnProArg 100
 Db 495 GCGGCGAAGAGCTGCTGTGTGATGCTGAACCGGACGCAAGCCCGCAAGCGCGGAGG 554
 QY 101 Arg 101
 Db 555 CGC 557

RESULT 5
 BM291800 715 bp mRNA linear EST 01-JUL-2002
 LOCUS BM291800
 DEFINITION EST574342 AvSG Amblyomma variegatum cDNA clone AVABC60 5' end, mRNA sequence.

ACCESSION BM291800 GI:21641767

VERSION BM291800

KEYWORDS EST.

SOURCE Amblyomma variegatum

ORGANISM Amblyomma variegatum

Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;

Parasitiformes; Ixodida; Ixodidae; Amblyomma.

1 (bases 1 to 715)

Nene, V., Lee, D., Quackenbush, J., Skilton, R., Mwaura, S.,

Gardner, M. J. and Bishop, R.

AvGI, an index of genes transcribed in the salivary glands of the

ixodid tick *Amblyomma variegatum*

Int. J. Parasitol. 32 (12), 1447-1456 (2002)

22281296

Contact: Vish Nene

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-610-5968

Fax: 301-838-0208

Email: nene@tigr.org

Seq primer: M13 reverse.

Location/Qualifiers

1..715

/organism="Amblyomma variegatum"

/mol_type="mRNA"

/db_xref="taxon:34610"

/clone="AVABC60"

/tissue_type="Salivary glands"

/dev_stage="Adult"

/lab_host="E.coli strain DH10B-Tona"

/clone_lib="AvSG"

/note="Vector: pCMV-SPORT6.1; Salivary glands were

dissected on day five after initiation of feeding. Total

RNA was prepared using acid guanidium thiocyanate-phenol-chloroform extraction. The cDNA library was custom prepared by Invitrogen Corporation. Briefly, a first strand cDNA was primed using oligo(dT) containing a NotI site. Size fractionated double stranded cDNA was ligated to EcoRV-NotI cleaved vector and electroporated into *E.coli*."

ORIGIN

Alignment Scores: 2.17e-44 Length: 715
 Pred. No.: 453.00 Matches: 79
 Score: 93.75% Conservative: 11
 Percent Similarity: 82.29% Mismatches: 6
 Best Local Similarity: 81.62% Indels: 0
 Query Match: 12 Gaps: 0
 DB:

US-09-844-353A-54 (1-103) x BM291800 (1-715)

QY 1 LysLysThrThrArgArgAsnAlaTrpGlyAsnMetSertYrAlaGluLeuIleThr 20
 Db 260 AAAAGAACTCGTGGCTCGCAACGCTGGGGCAACATGCTCTACGGGACCTGATTACG 319
 QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
 Db 320 CAAGCCATCCAAAGCGCGCGAGAAACGCGCTCTACAGATCTACGAGTGGATG 379
 QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
 Db 380 GTGCAGAACGTGCGCTATTTCAGAGACAGAGGCGACAGACAGCTCGCGCGTGGAG 439
 QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
 Db 440 AACTCTATCCGGCAACACCTGCTGTCACAGCAGATTTCATGCGAGTCCAGAACGAAGC 499
 QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAspAlaLysProGly 96
 Db 500 GCGGCGAAGAGCTGCTGTGTGATGCTGAACCCCGGACGCAAGCCCGGCG 547

RESULT 6

BM291800

LOCUS BM291800

DEFINITION EST574342 Sus Scrofa library (scac) Sus scrofa cDNA clone

scac00371.g.15 5prim, mRNA sequence.

ACCESSION BX673604

VERSION BX673604

KEYWORDS EST.

SOURCE Sus scrofa (pig)

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 593)

Bonnet, A., Tosser-Klopp, G., Benne, F., Cabau, C., Villegier, S.,

Soares, M., Bonaldo, F. and Hatey, F.

A Pig Normalised Multi-Tissue cDNA Library

Unpublished (2003)

Contact: Tosser-Klopp G

Genetique Animale

Institut National de la Recherche Agronomique

Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan

cedex, FRANCE

Tel: 33 (0) 5.61.28.51.14

Fax: 33 (0) 5.61.28.53.08

Email: tosser@toulouse.inra.fr

Clone distribution: AGENAE Resource centre, Francois PIUMI,

Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du

genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex,

FRANCE, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73

Sequence cleaned of vector, adaptor and repetitions. Contact us

at signenasupport@jouy.inra.fr to obtain the chromatogram of this

sequence.

Plate: 0037, row: g column: 15.

Location/Qualifiers

FEATURES

```

source
1. .593
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="scac00371.g.15"
/tissue_type="mixed"
/clone_lib="Sus Scrofa library (scac)"
/notes="Vector: pT7T3D-pac vector; tissues: adipose tissue,
brain, kidney, liver, muscle, ovary, testis, heart,
hypothalamus, pancreas, skin, spleen, thymus, placenta,
pituitary gland, seminal vesicle, small intestine,
uterus, adrenals, bulbo urethral gland, cerebral trunk,
epididymis, female gonad, gall-bladder, hippocampus,
large intestine, male gonad, melanocytes, stomach, udder"

ORIGIN

Alignment Scores:
Pred. No.: 5,35e-43 Length: 593
Score: 440.50 Matches: 79
Percent Similarity: 90.20% Conservative: 13
Best Local Similarity: 77.45% Mismatches: 9
Query Match: 79.37% Indels: 1
DB: 13 Gaps: 1

US-09-844-353A-54 (1-103) x BX673604 (1-593)

QY 1 LysLysThrThrThrArgArgAsnAlaTyrGlyAsnMetSerTyrAlaGluLeuLeuThr 20
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 74 AGGAATGCTCTCGCGCGGGAACGCTGCGGGAACCTGTCTTACGCCGATCTGATCACT 133

QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTyrMet 40
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 134 CGGCCCATCGAGAGTTCCCGGACAAACGGCTCACTCTGCCAGATCTATGATGGATG 193

QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTyrLys 60
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 194 GTGCGTGGCTGCGCTACTTCAAGGATAAGGGGACAGCAACAGTTCTTGGCGGCTGGAAG 253

QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAnGluGly 80
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 254 AACTCTATCCGGCACAACTGTCTCAAGGATAGGGGCGGTTCTATGCGGTCAGAACAGAGGGG 313

QY 81 AlaGlyLysSerSerTyrTyrPheValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 314 ACCGGCAGAGCTCTGTGGTGCATCATCAACCTTGACGGGGAAGAGCGGGAAGGCGCCC 373

QY 100 ArgArg 101
   |||||
Db 374 CGGCGG 379

RESULT 7
BM251446/c
LOCUS
DEFINITION
  BM251446 600 bp mRNA linear EST 17-DEC-2001
  (BOLT) Bos taurus cDNA 3', mRNA sequence.
ACCESSION
  BM251446
VERSION
  BM251446.1 GI:17887038
KEYWORDS
  EST.
SOURCE
  Bos taurus (cow)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
    Bovidae; Bovinae; Bos.
  1 (bases 1 to 600)
  Yao,J., Burton,J.L., Sipkovsky,S. and Coussens,P.M.
  Generation of EST and cDNA microarray resources for the study of
  bovine immunobiology
  Acta Vet. Scand. 42 (3), 391-406 (2001)
  JOURNAL
  MEDLINE
  21895187
  PUBMED
  11887399
  COMMENT
  Contact: Jianbo Yao
  Department of Animal Science
  Michigan State University

FEATURES
source
1. .600
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/sex="female"
/tissue_type="Blood"
/cell_type="peripheral blood leukocytes"
/dev_stage="mid-lactation"
/lab_host="DH10B"
/clone_lib="Normalized Bovine Total Leukocyte cDNA Library
(BOLT)"
/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

ORIGIN

Alignment Scores:
Pred. No.: 5,44e-43 Length: 600
Score: 440.50 Matches: 79
Percent Similarity: 90.20% Conservative: 13
Best Local Similarity: 77.45% Mismatches: 9
Query Match: 79.37% Indels: 1
DB: 12 Gaps: 1

US-09-844-353A-54 (1-103) x BM251446 (1-600)

QY 1 LysLysThrThrThrArgArgAsnAlaTyrGlyAsnMetSerTyrAlaGluLeuLeuThr 20
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 510 AGGAATGCTCTCGCGCGGGAACGCTGCGGGAACCTGTCTTACGCCGATCTGATCACT 451

QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTyrMet 40
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 450 CGGCGATCGAGACTCTCGACAAACGGCTCACTCTGCCAGATCTATGATGGATG 391

QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTyrLys 60
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 390 GTGCGTGGCTGCGCTACTTCAAGGATAAGGGGACAGCAACAGCTCTGCCGCTGGAAG 331

QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAnGluGly 80
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 330 AACTCTATCCGGCACAACTGTCTCAAGGATAGGGGCGGTTCTATGCGGTCAGAACAGAGG 271

QY 81 AlaGlyLysSerSerTyrTyrPheValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 270 ACCGGCAGAGCTCTGTGGTGCATCATCAACCTTGACGGGGAAGAGTGGCAAGCGCCC 211

QY 100 ArgArg 101
   |||||
Db 210 CGGCGG 205

RESULT 8
BE881329
LOCUS
DEFINITION
  BE881329 628 bp mRNA linear EST 20-OCT-2000
  mRNA sequence.
ACCESSION
  BE881329
VERSION
  BE881329.1 GI:10330105
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 628)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  JOURNAL
  COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: DCTD/DTP/Gazdar

```

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAN9679 row: b column: 05
 High quality sequence stop: 628.
 Location/Qualifiers

FEATURES

source
 1. 628
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3892516"
 /tissue_type="large cell carcinoma, undifferentiated"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 69"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."

ORIGIN

Alignment Scores:
 Pred. No.: 5.84e-43 Length: 628
 Score: 440.50 Matches: 79
 Percent Similarity: 90.20% Conservativeness: 13
 Best Local Similarity: 77.45% Mismatches: 9
 Query Match: 79.37% Indels: 1
 DB: 10 Gaps: 1

US-09-844-353A-54 (1-103) x BE881329 (1-628)

Qy 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
 Db 206 AGGAAATGTTCTCGCGCGGCAACGCTCTGGGGAACCTGTCTACGCGGACCTGATCACC 265

Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
 Db 266 CGCGCCATCGAGAGCTCCCGGACAAACGGCTCACTGTCTCCAGATCTACAGTGGAGT 325

Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
 Db 326 GTGGTTGCGTGGCTTCTTCAAGGATAAGGCGGACACACAGCTCTGCGCGGTGGAG 385

Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
 Db 386 AACTCCATCCGCGCACACCTGTCTACTGTCATAGTCGATTCAATCGCGGTCGAGATGAGGGA 445

Qy 81 AlaGlyLysSerSerTyrTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
 Db 446 ACTGCAAGAGCTCTTGTGGATCATCAACCTGTATGCGGGGAGAGCGGAAAGCCCC 505

Qy 100 ArgArg 101
 Db 506 CGGCGG 511

RESULT 9
 BC019532
 LOCUS
 DEFINITION Mus musculus, similar to forkhead box O3, clone IMAGE:4162687,
 mRNA.
 ACCESSION BC019532
 VERSION BC019532.1 GI:18044075
 KEYWORDS HTC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 998)
 Strausberg,R.
 Direct Submission
 Submitted (19-DEC-2001) National Institutes of Health, Mammalian

REMARK
COMMENT

Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCW-HGSC
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAC Plate: 37 Row: d Column: 23
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 9789950
 This clone has the following problem: frame shifted.

FEATURES

source

1. 998
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4162687"
 /tissue_type="Liver, normal. 5 month old male mouse."
 /clone_lib="NCI CGAP_Li9"
 /lab_host="DH10B"
 /note="vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:
 Pred. No.: 1.2e-42 Length: 998
 Score: 440.50 Matches: 79
 Percent Similarity: 90.20% Conservativeness: 13
 Best Local Similarity: 77.45% Mismatches: 9
 Query Match: 79.37% Indels: 1
 DB: 11 Gaps: 1

US-09-844-353A-54 (1-103) x BC019532 (1-998)

Qy 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
 Db 604 AGGAAATGTTCTCGCGCGGAAATGCTTGGGGGAACCTGTCTATGCGGACCTGATCACC 663

Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
 Db 664 CGCGCCATCGAGAGCTCCCGGACAAACGGCTCACTTGTCTCCAGATCTACAGTGGAGT 723

Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
 Db 724 GTGCGCTGTGTGCGCTTCTTCAAGGATAAGGCGGACAGACAGCTCTGCGGCTGGAG 783

Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
 Db 784 AACTCCATCCGCGCACACCTGTCTCTGACAGCGCTTCTGCGGCTTCAGATGAGGCG 843

Qy 81 AlaGlyLysSerSerTyrTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
 Db 844 ACGGCAAGAGCTCTTGTGGATCATCAACCCGATGGGGGAAGAGCGGAGGCCCCCC 903

Qy 100 ArgArg 101
 Db 904 CGGCGG 909

RESULT 10
 AK047413

LOCUS AK047413 2530 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length enriched library, Clone:B930059F01 product:forkhead box O3, full insert sequence.
ACCESSION AK047413 GI:26338755
VERSION AK047413.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 Carninci, P. and Hayashizaki, Y.
AUTHORS Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
AUTHORS Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, Y., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
AUTHORS RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
AUTHORS Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED 11076861
REFERENCE 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
AUTHORS Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12011000
REFERENCE 6 (bases 1 to 2530)
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiranoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kawai, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
AUTHORS Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
FEATURES
source
location/Qualifiers
1. 2530
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM, DB:B930059F01"
/db_xref="MGI:2412957"
/db_xref="taxon:10090"
/clone="B930059F01"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
284..2302
/note="unnamed protein product; forkhead box O3 (MGI:MGI:1890081, GB|NM_019740, evidence: BLASTN, 99%, match=2337)
putative"
/codon_start=1
/protein_id="BAC33049.1"
/db_xref="GI:26338756"
/translation="MAEAPASVPLSPLEVEIDPEPEPQSPRSCPTWLPQLOASP
APSGETAADSMIPEDDDDDGGGRASSAMVIGGVSYSTLGSGLLEDGAMLLAP
GGQDLSGSPASAGALSGGTQLOPQPLPOPGAGAGSGQPKCSSERNAGNLS
YADLITRAETSPDKELTSLQYEWVRCVPFKDGSNSAGKNSIRHLSHR
FWRVQNEGTSKSWIIPDGGKSPARRAVSMNSNYKTSRGRARAKKAQQA
PSADSPQLSGPGSTSGSDELDAWTFRSTNSNKTSTVSGRLSTLASELDD
VDDGDLSPMLYSSASLSPVYKPSGLGTSFGSTFVFPSSLSLRQSPMOTI
IALPSPGPGGLMORGSSFTVAKSGLSGLSTFVFPSSLSLRQSPMOTI
QENRATFSSVSHYGNQTLQDILLASLSHSDVMQSDPLMSQASTAVSAQNRNV
MERNDPMFSFAQPTGSLVQNQLLHHQHTQALGSGRALSNSVSNGLSGSLGS
AKHQQSPASQSMQILSSGLSLSASANLPMVGHDKFPDLDLMDMFGSLCDME
SIIRSELMDADGLDFNLSLISTQNVVGLNVGNFTGAKQASSQSWVPG"
2507..2512
/note="putative"
2530
/note="putative"
ORIGIN
Alignment Scores:
Pred. No.: 5.1e-42 Length: 2530
Score: 440.50 Matches: 79
Percent Similarity: 90.20% Conservative: 13
Best Local Similarity: 77.45% Mismatches: 9
Query Match: 79.37% Indels: 1
DB: 11 Gaps: 1
US-09-844-353A-54 (1-103) x AK047413 (1-2530)
QY 1 LysLysThrThrThrArgArgGlnalaTrpGlyAsnMetSerTyAlaGluLeuThr 20
722 AGGAATGCTCTCCGCGGGAATCCTGGGGAAACCTGTCTCTATGCCACCTGATCACC 781
Db ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyGluTrpMet 40
QY 782 CGCGCATCGAGAGTCCCGGACAAAGCGTCACTTTGTCCAGATCACCAGTGGGATG 841
Db ValGlnAsnValProTyPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
QY 842 GTGCGCTGTGTGCCCTACTTCAAGGATAAGGCGACAGCAACAGCTCTCGGGCTGGAG 901
Db AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
QY 902 AACTCATCCGCGCAACACTGTCCCTGCACACCGCTTTCATGCGCGTTTCAGAAATGAAGC 961
Db AlaGlyLysSerSerTipTipValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
QY 962 ACGGCAAGAGCTCTTGTGTGATCATCAACCCGATGGGGAAGAGCGGGAGGCCCCC 1021

QY 100 ArgArg 101
 Db 1022 CGCGG 1027

RESULT 11
 AV616467
 LOCUS 448 bp mRNA linear EST 28-NOV-2001
 DEFINITION Bos taurus ovary fetus Bos taurus cDNA clone ELOV005E02
 3', mRNA sequence.
 ACCESSION AV616467
 VERSION AV616467.1 GI:9752137
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 448)
 AUTHORS Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
 and Sugimoto,Y.
 TITLE Establishment of a high throughput EST sequencing system using
 poly(A) tail-removed cDNA libraries and determination of 36,000
 bovine ESTs
 JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
 MEDLINE 21570554
 PUBMED 11713328
 COMMENT Contact: Yoshikazu Sugimoto
 Animal Genetics Division
 Shirakawa Institute of Animal Genetics
 Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
 Tel: 81-248-25-5641
 Fax: 81-248-25-5725
 Email: kazusugie@ocn.ocn.ne.jp
 Single pass sequencing.
 This clone was obtained from a polyA-deleted cDNA library.

FEATURES
 source
 1..448
 Location/Qualifiers
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /clone="ELOV005E02"
 /tissue_type="ovary"
 /dev_stage="fetus"
 /lab_host="DH10B"
 /clone_lib="Bos taurus ovary fetus"
 /note="Vector: pZLI; Site1: SalI; Site2: NotI; Poly A
 was deleted from a NotI site"

ORIGIN
 Alignment Scores:
 Pred. No.: 1,318-41 Length: 448
 Score: 427.50 Matches: 75
 Percent Similarity: 91.18% Conservative: 18
 Best Local Similarity: 73.53% Mismatches: 8
 Query Match: 77.03% Indels: 1
 DB: 9 Gaps: 1

US-09-844-353A-54 (1-103) x AV616467 (1-448)

QY 1 LysLysThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuLeuThr 20
 Db 34 AAGAGCAGTCTGCTCCCGCGCAACGCTGGGCAACCTGCTTACGCGGACCTCATCACC 93
 QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
 Db 94 AAGGCCATCGAGAGTCCGCCGAGAGGAGGACTCAGCTGTCGCATTTACGAGTGGAG 153
 QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
 Db 154 GTCAAGAGCGTGCCTTACTTCAAGGATAGGGGACGACGACAGCTCAGCGGCTGGAG 213
 QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80

Db 214 AATTCAATTCGTATATCTGTCCTCCACAGCAAGTTTCATCCCGTGCAGATGAAGGA 273

QY 81 AlagLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
 Db 274 ACTGGAAAAGCTCTGGTGGATGCTCATCCAGAGGAGGAGAGTGGGAAATCCCC 333

QY 100 ArgArg 101
 Db 334 AGGAGA 339

RESULT 12
 AW204560
 LOCUS 459 bp mRNA linear EST 02-DEC-1999
 DEFINITION UI-H-B11-aec-d-05-0-UI.s1 NCI CGAP_Sub3 Homo sapiens cDNA clone
 IMAGE:2718776 3', mRNA sequence.
 ACCESSION AW204560
 VERSION AW204560.1 GI:6504032
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 459)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgap@remail.nih.gov
 Oligo-dt track not found, Not I site shown in beginning of sequence
 is likely internal to the message. cDNA Library preparation: M.B.
 Soares Lab Clone distribution: NCI-CGAP clone distribution
 information can be found through the I.M.A.G.E. Consortium/ILNL at:
 www-bio.llnl.gov/bbrp/image/image.html The following repetitive
 elements were found in this cDNA sequence: 64-150,
 >(CGG)n#Simple_repeat
 Seq primer: M13 Forward
 POLYA=No.

FEATURES
 source
 1..459
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2718776"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP Sub3"
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NCI CGAP Sub3 library is a subtracted library derived from
 the NCI CGAP Sub1 library, which is a subtracted library
 derived from B1. B1 constitutes a mixture of 21
 normalized or subtracted NCI CGAP libraries:
 NCI CGAP Co4, NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP Co10,
 NCI CGAP Co16, NCI CGAP Kid5, NCI CGAP Kid12,
 NCI CGAP Kid3, NCI CGAP Kid11, NCI CGAP Lym2,
 NCI CGAP Br2, NCI CGAP Co8, NCI CGAP CLL1, NCI CGAP Lei2,
 NCI CGAP Brn23, NCI CGAP Lu5, NCI CGAP Lu24,
 NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6,
 NCI CGAP Brn25. These 21 libraries were pooled and a
 single-stranded DNA preparation of the resulting mixture
 was used as a tracer in a subtractive hybridization with
 a driver whose composition is detailed below:
 NCI CGAP Kid3 pool 1 LLAM 3334-3337, 3682-3683,
 3798-3803 (IMAGE CloneIDs 1322376-1323911,
 1456008-1456775, 1500552-1502855); NCI CGAP Kid5 pool 1
 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIDs
 132912-1325831, 1471368-1472903, 1492104-1493255);
 NCI CGAP Lys pool 1 LLAM 3575-3582, 3851-3854 (IMAGE
 CloneIDs 1414920-1417991, 1520904-1522439); NCI CGAP GC4
 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
 CloneIDs 1257096-1258631, 1469064-1470983,
 1475592-1476743); NCI CGAP Pr22 pool 1 LLAM 2457-2459,
 2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,

Db	334	AATTCAATTCGTATAATCTCTCCCTACACGAAGTTCATTTCGTGCAGAAATCAAGGA	393
Qy	81	AlaGlyLysSerSerTyrTyrValIleAsnProAsp---AlaLysProGlyMetAsnPro	99
Db	394	ACTGGAAAAGTCTCTTCGGTGGATGCTCAATCCAGAGGTCGCAAGAGCGGGAAATCTCCT	453
Qy	100	ArgArg 101	
Db	454	AGGAGA 459	

Search completed: July 29, 2004, 12:13:31
Job time : 2496 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 29, 2004, 09:52:26 ; Search time 3081 Seconds
(without alignments)
1448.988 Million cell updates/sec

Title: US-09-844-353A-54

Perfect score: 555

Sequence: 1 KKTTRRNAGNWSYAEILIT.....SSWWINPDAGKGNPRTR 103

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlpl
-Q=/cgm2_1/USPTO_spool_p/US09844353/runat_23072004_084913_28237/app_query.fasta_1.263
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCI=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09844353.@CGN_1_7406.@runat_23072004_084913_28237 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.cm.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	555	100.0	3499	6	AR148563 Sequence
2	555	100.0	3499	6	BD193386 Therapeut
3	549	98.9	2565	3	AF020343 Caenorhab
4	549	98.9	2579	3	AF020342 Caenorhab
5	549	98.9	3035	3	AF032112 Caenorhab
6	472	85.0	2589	3	AY281749 Strongylo
7	450	81.1	3513	3	AF426831 Drosophil
8	450	81.1	3934	3	AY089542 Drosophil
9	440.5	79.4	2685	9	AK092357 Homo sapi
10	440.5	79.4	2870	9	AK122861 Homo sapi
11	440.5	79.4	2889	10	AF114259 Mus muscu
12	440.5	79.4	3183	9	AF032886 Homo sapi
13	440.5	79.4	3300	9	BC020227 Homo sapi
14	440.5	79.4	3300	9	BC021224 Homo sapi
15	435.5	78.5	5742	5	AF114261 Gallus ga
16	433.5	78.1	2740	5	AY040320 Xiphophor
17	427.5	77.0	1968	9	BT007455 Homo sapi
18	427.5	77.0	2162	4	AY094061 Sus scrof
19	427.5	77.0	2413	9	BC021981 Homo sapi
20	427.5	77.0	2503	10	AY255525 Spermothi
21	427.5	77.0	3421	9	HSU02310 Human fork
22	427.5	77.0	5723	6	AX587664 Sequence
23	427.5	77.0	5723	9	AF032885 Homo sapi
24	427.5	77.0	5769	6	AX779990 Sequence
25	426.5	76.8	2103	10	AF126056 Mus muscu
26	426.5	76.8	2460	10	AF114258 Mus muscu
27	426.5	76.8	4945	10	MMU252157 Mus muscu
28	425.5	76.7	4547	5	AF114262 Danio rer
29	419.5	75.6	2528	10	MMU571690 Mus muscu
30	416.5	75.0	3244	9	AF032887 Homo sapi
31	416.5	75.0	171978	9	AC026271 Homo sapi
32	413.5	74.5	1733	10	AB032770 Mus muscu
33	413.5	74.5	1993	10	AF114260 Mus muscu
34	410.5	74.0	1073	6	AR242549 Sequence
35	410.5	74.0	3394	6	AR338823 Sequence
36	406.5	73.2	3171	6	AX821959 Sequence
37	406.5	73.2	3171	9	HSAFX
38	388.5	70.0	1722	9	HSAB1589
39	378.5	68.2	2704	6	AR148564 Sequence
40	378.5	68.2	2704	6	BD193387 Therapeut
41	378	68.1	900	5	AF064828 Danio rer
42	372.5	67.1	2345	3	AF020344 Caenorhab
43	356	64.1	1873	3	AY281750 Strongylo
44	339.5	61.2	1000	10	AF247812 Rattus no
45	335.5	60.5	1353	6	AX665875 Sequence

ALIGNMENTS

RESULT 1

AR148563
 LOCUS AR148563 3499 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 43 from patent US 6225120.
 ACCESSION AR148563
 VERSION AR148563.1 GI:15112653
 KEYWORDS
 SOURCE Unknown.
 ORGANISM
 Unclassified.
 REFERENCE 1 (bases 1 to 3499)
 AUTHORS Ruvkun,G., Kimura,K., Patterson,G., Ogg,S., Paradis,S., Tissenbaum,H., Morris,J. and Koweek,A.
 TITLE Therapeutic and diagnostic tools for impaired glucose tolerance conditions
 JOURNAL Patent: US 6225120-A 43 01-MAY-2001;
 FEATURES Location/Qualifiers
 source 1..3499
 /organism="unknown"
 /mol_type="unassigned DNA"

ORIGIN
 Alignment Scores:
 Pred. No.: 4.1e-52 Length: 3499
 Score: 555.00 Matches: 103
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-844-353A-54 (1-103) x AR148563 (1-3499)
 QY 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyAlaGluLeuIleThr 20
 DB 727 AAAAAGACACACGACGCGCTTGGGAAATATGTCATATGCTGAATCTATCACT 786
 QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyGluTrpMet 40
 DB 787 ACAGCCATTATGGCTAGTCCAGAGAAACGGTTAACTCTTGACAAAGTTTACGAATGGATG 846
 QY 41 ValGlnAsnValProTyPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
 DB 847 GTCCAGAAATGTTCCATACTTCCAGGATAAGGAGATTCCGAAGTTCAGCTGGATGGAAAG 906
 QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
 DB 907 AACTCGATCCGTCACAACTGCTCTTCACTTCGTTTCATGCGAATTCAGAAATCGAAGA 966
 QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAspAlaLysProGlyMetAsnProArg 100
 DB 967 GCCCGAAAGAGCTCGTGGTGGGTATTATATCCAGATGCAAGCCAGGAATGAATCCACGG 1026
 QY 101 ArgThrArg 103
 DB 1027 CGTACACGT 1035

RESULT 2
 BD193386
 LOCUS BD193386 3499 bp DNA linear PAT 17-JUL-2003
 DEFINITION Therapeutic and diagnostic tools for impaired glucose tolerance conditions.
 ACCESSION BD193386
 VERSION BD193386.1 GI:33003125
 KEYWORDS JP 2002511747-A/15.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 3499)
 AUTHORS Ruvkun,G., Kimura,K., Patterson,G., Ogg,S., Paradis,S., Tissenbaum,H., Morris,J., Koweek,A. and Pierce,S.
 TITLE Therapeutic and diagnostic tools for impaired glucose tolerance conditions
 JOURNAL Patent: JP 2002511747-A 15 16-APR-2002;

COMMENT

THE GENERAL HOSPITAL CORP
 OS Caenorhabditis elegans
 PN JP 2002511747-A/15
 PD 16-APR-2002
 PF 15-MAY-1998 JP 1998549639
 PR 15-MAY-1997 US 08/857076, 07-JUL-1997 US 08/888534 PI
 GARY RUVKUN, KOTARO KIMURA, GARTH PATTERSON, SCOTT OGG, SUZANNE PI
 PARADIS.
 PI HEIDI TISSENBAUM, JASON MORRIS, ALLISON KOWEEK, SARAH PIERCE PC
 A61K49/00, C12N5/06, C07H21/04
 CC Therapeutic and diagnostic tools for impaired glucose CC
 tolerance conditions
 FH Key Location/Qualifiers
 FT source 1..3499
 /organism="Caenorhabditis elegans".

FEATURES

source 1..3499
 /organism="Caenorhabditis elegans"
 /mol_type="genomic DNA"
 /db_xref="taxon:6239"

ORIGIN

Alignment Scores:
 Pred. No.: 4.1e-52 Length: 3499
 Score: 555.00 Matches: 103
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-844-353A-54 (1-103) x BD193386 (1-3499)

QY 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyAlaGluLeuIleThr 20
 DB 727 AAAAAGACACACGACGCGCTTGGGAAATATGTCATATGCTGAATCTATCACT 786
 QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyGluTrpMet 40
 DB 787 ACAGCCATTATGGCTAGTCCAGAGAAACGGTTAACTCTTGACAAAGTTTACGAATGGATG 846
 QY 41 ValGlnAsnValProTyPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
 DB 847 GTCCAGAAATGTTCCATACTTCCAGGATAAGGAGATTCCGAAGTTCAGCTGGATGGAAAG 906
 QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
 DB 907 AACTCGATCCGTCACAACTGCTCTTCACTTCGTTTCATGCGAATTCAGAAATCGAAGA 966
 QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAspAlaLysProGlyMetAsnProArg 100
 DB 967 GCCCGAAAGAGCTCGTGGTGGGTATTATATCCAGATGCAAGCCAGGAATGAATCCACGG 1026
 QY 101 ArgThrArg 103
 DB 1027 CGTACACGT 1035

RESULT 3

AF020343
 LOCUS AF020343 2565 bp mRNA linear INV 15-NOV-1997
 DEFINITION Caenorhabditis elegans fork head-related transcription factor DAF-16a2 (daf-16) mRNA, complete cds.
 ACCESSION AF020343
 VERSION AF020343.1 GI:2618978
 KEYWORDS
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 2565)
 AUTHORS Ogg,S., Paradis,S., Gottlieb,S., Patterson,G.I., Lee,L., Tissenbaum,H.A. and Ruvkun,G.
 TITLE The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in C. elegans

JOURNAL Nature 389 (6654), 994-999 (1997)
 MEDLINE 98013175
 PUBMED 9353126
 REFERENCE 2 (bases 1 to 2565)
 AUTHORS Ogg, S., Paradis, S. and Ruvkun, G.
 TITLE Direct Submission
 JOURNAL Submitted (21-AUG-1997) Molecular Biology, MGH, 50 Blossom St., Boston, MA 02114, USA

FEATURES
 source Location/Qualifiers
 1..2565
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /chromosome="1"
 /map="1; between mgp45 and mgp49"
 1..2565
 /gene="daf-16"
 /note="null phenotype is suppression of daf-2"
 297..1823
 /gene="daf-16"
 /codon_start=1
 /product="fork head-related transcription factor DAF-16a2"
 /protein_id="AAB84391.1"
 /db_xref="GI:2618979"
 /translation="MMEMLVDOGTDASSASTSTSSVRFGADTFMNTDDVMNDMM
 EPIPRDRCNTWMPRRPQLEPLNSPIIHEQIPEEDADLYGSNEQCQLGAGSSNGST
 AMLHTPDGNSHOTSPSESSPDIVSGKTTTRNAGNMSYAEILITTAISGPE
 KRULTAQYEMWQNVFYFRDKGDSNAGWKNSIRHNSLHSFRMRIQNEGAGKSW
 WVINPAKPGRNPRTERSENTIETTTKAQLEKSRGAKKRIKERALMGLSHLTNGN
 SIAGSIQTI SHDLYDDSMQCAFNDVPSPFRPTQSNLSIPGSSSRVSPAIGSDIYDD
 LEPPSWGVESVPAIPSDIVDRTQMDRIDATTHIGGVQIKQESKPIKTEPIAPPSYHE
 LNSVRGSCAQNPILRNPIVSTNFKPMLPGAYGNQNGGIIPIINLWSTNSSPLEGI
 QSGCGIVAQHTVASSALPIDLENLITLDPQLMDTMDVDLIRHELSQLGGQHIHFDL
 "

gene
 1..2565
 /note="null phenotype is suppression of daf-2"
 297..1823
 /gene="daf-16"
 /codon_start=1
 /product="fork head-related transcription factor DAF-16a1"
 /protein_id="AAB84390.1"
 /db_xref="GI:2618977"
 /translation="MMEMLVDOGTDASSASTSTSSVRFGADTFMNTDDVMNDMM
 EPIPRDRCNTWMPRRPQLEPLNSPIIHEQIPEEDADLYGSNEQCQLGAGSSNGST
 AMLHTPDGNSHOTSPSESSPDIVSGKTTTRNAGNMSYAEILITTAISGPE
 KRULTAQYEMWQNVFYFRDKGDSNAGWKNSIRHNSLHSFRMRIQNEGAGKSW
 WVINPAKPGRNPRTERSENTIETTTKAQLEKSRGAKKRIKERALMGLSHLTNGN
 SIAGSIQTI SHDLYDDSMQCAFNDVPSPFRPTQSNLSIPGSSSRVSPAIGSDIYDD
 LEPPSWGVESVPAIPSDIVDRTQMDRIDATTHIGGVQIKQESKPIKTEPIAPPSYHE
 LNSVRGSCAQNPILRNPIVSTNFKPMLPGAYGNQNGGIIPIINLWSTNSSPLEGI
 QSGCGIVAQHTVASSALPIDLENLITLDPQLMDTMDVDLIRHELSQLGGQHIHF
 DL"

CDS
 297..1823
 /gene="daf-16"
 /codon_start=1
 /product="fork head-related transcription factor DAF-16a1"
 /protein_id="AAB84390.1"
 /db_xref="GI:2618977"
 /translation="MMEMLVDOGTDASSASTSTSSVRFGADTFMNTDDVMNDMM
 EPIPRDRCNTWMPRRPQLEPLNSPIIHEQIPEEDADLYGSNEQCQLGAGSSNGST
 AMLHTPDGNSHOTSPSESSPDIVSGKTTTRNAGNMSYAEILITTAISGPE
 KRULTAQYEMWQNVFYFRDKGDSNAGWKNSIRHNSLHSFRMRIQNEGAGKSW
 WVINPAKPGRNPRTERSENTIETTTKAQLEKSRGAKKRIKERALMGLSHLTNGN
 SIAGSIQTI SHDLYDDSMQCAFNDVPSPFRPTQSNLSIPGSSSRVSPAIGSDIYDD
 LEPPSWGVESVPAIPSDIVDRTQMDRIDATTHIGGVQIKQESKPIKTEPIAPPSYHE
 LNSVRGSCAQNPILRNPIVSTNFKPMLPGAYGNQNGGIIPIINLWSTNSSPLEGI
 QSGCGIVAQHTVASSALPIDLENLITLDPQLMDTMDVDLIRHELSQLGGQHIHF
 DL"

ORIGIN
 Alignment Scores:
 Pred. No.: 1 36e-51 Length: 2565
 Score: 549.00 Matches: 102
 Percent Similarity: 99.03% Conservative: 0
 Best Local Similarity: 99.03% Mismatches: 1
 Query Match: 98.92% Indels: 0
 DB: 3 Gaps: 0

US-09-844-353A-54 (1-103) x AF020343 (1-2565)

QY 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuLeuThr 20
 Db 693 AAAAAACACACAGCAGCAGGACGCTTGGGAAATATGTCATATGCTGACTTATCACT 752

QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
 Db 753 ACAGCCATTATGCTAGTCAGAGAACCGTTAACTCTTGCAACAGTTTACGAATGATG 812

QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTTrpLys 60
 Db 813 GTCCAGATGTTCCATCTTACAGGATACAGGAGATTCGACAGTTCAGTCGATGGAAG 872

QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
 Db 873 AACTCGATCGTCACAAATGCTCTCTTCATCTTCGTTTCATGCGATTTCAGATGAGGA 932

QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAspAlaLysProGlyMetAsnProArg 100
 Db 933 GCCGGAAGAGCTCGTGGTGGGTATTATATCCAGATGCAAGCCAGGAAGGATCCACGG 992

QY 101 ArgThrArg 103
 Db 993 CGTACACGT 1001

RESULT 4
 AF020342

LOCUS AF020342 2579 bp mRNA linear INV 13-NOV-1997
 DEFINITION Caenorhabditis elegans fork head-related transcription factor
 DAF-16a1 (daf-16) mRNA, complete cds.
 ACCESSION AF020342
 VERSION AF020342.1 GI:2618976
 KEYWORDS
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 2579)
 AUTHORS Ogg, S., Paradis, S., Gottlieb, S., Patterson, G.I., Lee, L.,
 Tissenbaum, H.A. and Ruvkun, G.
 TITLE The Fork head transcription factor DAF-16 transduces insulin-like
 metabolic and longevity signals in C. elegans
 JOURNAL Nature 389 (6654), 994-999 (1997)
 MEDLINE 98013175
 PUBMED 9353126
 REFERENCE 2 (bases 1 to 2579)
 AUTHORS Ogg, S., Paradis, S. and Ruvkun, G.
 TITLE Direct Submission
 JOURNAL Submitted (21-AUG-1997) Molecular Biology, MGH, 50 Blossom St., Boston, MA 02114, USA

FEATURES
 source Location/Qualifiers
 1..2579
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /chromosome="1"
 /map="1; between mgp45 and mgp49"
 1..2579
 /gene="daf-16"
 /note="null phenotype is suppression of daf-2"
 299..1831
 /gene="daf-16"
 /codon_start=1
 /product="fork head-related transcription factor DAF-16a1"
 /protein_id="AAB84390.1"
 /db_xref="GI:2618977"
 /translation="MMEMLVDOGTDASSASTSTSSVRFGADTFMNTDDVMNDMM
 EPIPRDRCNTWMPRRPQLEPLNSPIIHEQIPEEDADLYGSNEQCQLGAGSSNGST
 AMLHTPDGNSHOTSPSESSPDIVSGKTTTRNAGNMSYAEILITTAISGPE
 KRULTAQYEMWQNVFYFRDKGDSNAGWKNSIRHNSLHSFRMRIQNEGAGKSW
 WVINPAKPGRNPRTERSENTIETTTKAQLEKSRGAKKRIKERALMGLSHLTNGN
 SIAGSIQTI SHDLYDDSMQCAFNDVPSPFRPTQSNLSIPGSSSRVSPAIGSDIYDD
 LEPPSWGVESVPAIPSDIVDRTQMDRIDATTHIGGVQIKQESKPIKTEPIAPPSYHE
 LNSVRGSCAQNPILRNPIVSTNFKPMLPGAYGNQNGGIIPIINLWSTNSSPLEGI
 QSGCGIVAQHTVASSALPIDLENLITLDPQLMDTMDVDLIRHELSQLGGQHIHF
 DL"

gene
 1..2579
 /note="null phenotype is suppression of daf-2"
 299..1831
 /gene="daf-16"
 /codon_start=1
 /product="fork head-related transcription factor DAF-16a1"
 /protein_id="AAB84390.1"
 /db_xref="GI:2618977"
 /translation="MMEMLVDOGTDASSASTSTSSVRFGADTFMNTDDVMNDMM
 EPIPRDRCNTWMPRRPQLEPLNSPIIHEQIPEEDADLYGSNEQCQLGAGSSNGST
 AMLHTPDGNSHOTSPSESSPDIVSGKTTTRNAGNMSYAEILITTAISGPE
 KRULTAQYEMWQNVFYFRDKGDSNAGWKNSIRHNSLHSFRMRIQNEGAGKSW
 WVINPAKPGRNPRTERSENTIETTTKAQLEKSRGAKKRIKERALMGLSHLTNGN
 SIAGSIQTI SHDLYDDSMQCAFNDVPSPFRPTQSNLSIPGSSSRVSPAIGSDIYDD
 LEPPSWGVESVPAIPSDIVDRTQMDRIDATTHIGGVQIKQESKPIKTEPIAPPSYHE
 LNSVRGSCAQNPILRNPIVSTNFKPMLPGAYGNQNGGIIPIINLWSTNSSPLEGI
 QSGCGIVAQHTVASSALPIDLENLITLDPQLMDTMDVDLIRHELSQLGGQHIHF
 DL"

CDS
 299..1831
 /gene="daf-16"
 /codon_start=1
 /product="fork head-related transcription factor DAF-16a1"
 /protein_id="AAB84390.1"
 /db_xref="GI:2618977"
 /translation="MMEMLVDOGTDASSASTSTSSVRFGADTFMNTDDVMNDMM
 EPIPRDRCNTWMPRRPQLEPLNSPIIHEQIPEEDADLYGSNEQCQLGAGSSNGST
 AMLHTPDGNSHOTSPSESSPDIVSGKTTTRNAGNMSYAEILITTAISGPE
 KRULTAQYEMWQNVFYFRDKGDSNAGWKNSIRHNSLHSFRMRIQNEGAGKSW
 WVINPAKPGRNPRTERSENTIETTTKAQLEKSRGAKKRIKERALMGLSHLTNGN
 SIAGSIQTI SHDLYDDSMQCAFNDVPSPFRPTQSNLSIPGSSSRVSPAIGSDIYDD
 LEPPSWGVESVPAIPSDIVDRTQMDRIDATTHIGGVQIKQESKPIKTEPIAPPSYHE
 LNSVRGSCAQNPILRNPIVSTNFKPMLPGAYGNQNGGIIPIINLWSTNSSPLEGI
 QSGCGIVAQHTVASSALPIDLENLITLDPQLMDTMDVDLIRHELSQLGGQHIHF
 DL"

ORIGIN
 Alignment Scores:
 Pred. No.: 1 37e-51 Length: 2579
 Score: 549.00 Matches: 102
 Percent Similarity: 99.03% Conservative: 0
 Best Local Similarity: 99.03% Mismatches: 1
 Query Match: 98.92% Indels: 0
 DB: 3 Gaps: 0

US-09-844-353A-54 (1-103) x AF020342 (1-2579)

QY 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuLeuThr 20
 Db 701 AAAAAACACACACAGCAGGACGCTTGGGAAATATGTCATATGCTGACTTATCACT 760

QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
 Db 761 ACAGCCATTATGCTAGTCAGAGAACCGTTAACTCTTGCAACAGTTTACGAATGATG 820

QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
 Db 821 GTCCAGATGTTCCATCTTACAGGATACAGGAGATTCGACAGTTCAGTCGATGGAAG 880

```

Qy 61 AnSerlleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluCly 80
Db 881 AACTCGATCCGTCACAAATCTGCTCTTCATTCGTTTCATGCGAAATTCAGAAATGAAGGA 940
Qy 81 AlaGlyLysSerSerTTPValIleAsnProAspAlaLysProGlyMetAsnProArg 100
Db 941 GCCGGAAGAGCTCGTGTGGGTATTAAATCCAGATGCAAGCCAGAGGAAATCCACGG 1000
Qy 101 ArgThrArg 103
Db 1001 CGTACACGT 1009

RESULT 5
AF032112 3035 bp mRNA linear INV 19-NOV-1997
LOCUS daf-16: An HNF-3/forkhead family member that can function to double
DEFINITION Caenorhabditis elegans DAF-16 (daf-16) mRNA, complete cds.
ACCESSION AF032112
VERSION AF032112.1 GI:2623942
KEYWORDS Caenorhabditis elegans
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 3035)
Lin, K., Dorman, J.B., Rodan, A. and Kenyon, C.
the life-span of Caenorhabditis elegans
Science 278 (5341), 1319-1322 (1997)
JOURNAL 98028757
MEDLINE 9360933
PUBMED
REFERENCE 2 (bases 1 to 3035)
AUTHORS Lin, K., Dorman, J.B., Rodan, A. and Kenyon, C.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-1997) Biochemistry and Biophysics, UCSF, 513
Panassus Ave., San Francisco, CA 94143-0554, USA
FEATURES
source
1..3035
Location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
1..3035
/gene="daf-16"
/Note="defective dauer formation; suppresses Age phenotype
of age-1, daf-2"
334..1866
/gene="daf-16"
/Note="HNF-3/forkhead homolog"
/codon_start=1
/product="DAF-16"
/protein_id="AAC47803.1"
/db_xref="GI:2623943"
/translation="WMEMLVDOGTASSASTSTSVSRFGADTFMNTDPDDVMNDMM
RPIPRDNTWMPRPOLPEPLNSPIIHEQIPEDADLYSGNOCGOLGASSNGST
AMLPDSSNSHOTSPFDPFMSSEPDITVSGKTTTRNAGNWSYAEILITAMAS
PEKRLTAQVIEWMVNPYFRDKGSSNSAGKNSIRHNLHLSRFRRIQNEGAKS
SNWVNPDAKPNRPRTRENTTETTTKALEKSRGAKKRIKERALMDGLSHSTLN
GNSIAGSQTTSHDLYDDSMQAFDPVSPFRPTQNSLSPGSSRVSPIAGSDIY
DLEFPWSVGEVPAIPSDIVDRTPQMRIDATHTGGVQIKQESKPIKTEPIAPPYS
HELNSVRGSCAONPLIRNPVPTSTNFKMPLPGAYGNQGGITPINWLSTNSPLP
GLQSGIVAAQHTVASSNALPIDLENLITLPDQPLMDTMDVDALIRHELSSQAGQHIHF
DL"

ORIGIN
Alignment Scores:
Pred. No.: 1..65E-51 Length: 3035
Score: 549.00 Matches: 102
Percent Similarity: 99.03% Conservative: 0
Best Local Similarity: 99.03% Mismatches: 1
Query Match: 98.92% Indels: 0
DB: 3 Gaps: 0

```

```

US-09-844-353A-54 (1-103) x AF032112 (1-3035)
Qy 1 LysLysThrThrArgArgAsnAlaTTPGlyAsnMetSerTyrAlaGluLeuIleThr 20
Db 736 AAAAAGACAACGACGACAGACGCGTGGGAAATATGTCATGCTGAACCTATCACT 795
Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
Db 796 ACAGCCATTTGGCTAGTCCAGAAACGGTTAACTCTTGACAAATTACGAATGGATG 855
Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db 856 GTCCAGAAATGTTCCATACCTTCAGGATAAAGGAGATTGCAACAGTTCAGCTGGTGAAG 915
Qy 61 AnSerlleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluCly 80
Db 916 AACTCGATCCGTCACAAATCTGCTCTTCATTCGTTTCATGCGAAATTCAGAAATGAAGGA 975
Qy 81 AlaGlyLysSerSerTTPValIleAsnProAspAlaLysProGlyMetAsnProArg 100
Db 976 GCCGGAAGAGCTCGTGTGGGTATTAAATCCAGATGCAAGCCAGAGGAAATCCACGG 1035
Qy 101 ArgThrArg 103
Db 1036 CGTACACGT 1044

RESULT 6
AF0321749 2589 bp mRNA linear INV 31-OCT-2003
LOCUS Strongyloides stercoralis forkhead transcription factor 1 isoform a
DEFINITION (fkf-1) mRNA, complete cds; alternatively spliced.
ACCESSION AY281749
VERSION AY281749.1 GI:33621211
KEYWORDS Strongyloides stercoralis
SOURCE Strongyloides stercoralis
ORGANISM Strongyloides stercoralis
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimoidea; Strongyloidea; Strongyloides.
1 (bases 1 to 2589)
Massey, H.C. Jr., Nishi, M., Chaudhary, K., Pakpour, N. and Lok, J.B.
Structure and developmental expression of Strongyloides stercoralis
fkf-1, a proposed ortholog of daf-16 in Caenorhabditis elegans
Int. J. Parasitol. 33 (13), 1537-1544 (2003)
JOURNAL 22935308
MEDLINE 14572516
PUBMED
REFERENCE 2 (bases 1 to 2589)
AUTHORS Massey, H.C. Jr., Nishi, M., Chaudhary, K., Pakpour, N. and Lok, J.B.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2003) Pathobiology, Univ. of Pennsylvania
Veterinary School, 430 Rosenthal, 3800 Spruce Street, Philadelphia,
PA 19104-6008, USA
FEATURES
source
1..2589
Location/Qualifiers
/organism="Strongyloides stercoralis"
/mol_type="mRNA"
/strain="UPD"
/isolation_source="dog"
/db_xref="taxon:6248"
/dev_stage="infective third stage larvae (L3i)"
/Note="University of Pennsylvania strain"
1..2589
/gene="fkf-1"
/Note="similar to Caenorhabditis elegans daf-16"
misc_feature 1
/gene="fkf-1"
/Note="splice leader acceptor site; determined by binding
of 5' RACE anchor primer with Caenorhabditis elegans S11
splice leader sequence"
211..2436
/gene="fkf-1"
/Note="FKTF-1a; DNA-binding; alternatively spliced"
/codon_start=1
CDS

```



```
Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTyrMet 40
Db 715 CATGCCATTGATCGGCCACGACAAACGATTTGACTGATGATTTACGATGGATG 774
Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTyrLys 60
Db 775 GTCCAGATATGCCATATTTCAAGACACAGGCGCATTCGATAGCATGCGCGATGGAAG 834
Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db 835 AACTCCATACGTCAATCTGCTGTCACACCGCTTTATGAGGGTCCAAACAGAGGC 894
Qy 81 AlaGlyLysSerSerTyrTyrValIleAsnProAspAlaLysProGlyMetAsnProArg 100
Db 895 ACCGGCAAGTCATCTGCTGATGCTCAACCCGAGGCGCAAGCGCGCAAGTCTGTGGCG 954
Qy 101 Arg 101
Db 955 CGC 957

RESULT 8
AY089542
LOCUS Drosophila melanogaster LD19191 full insert cDNA.
ACCESSION AY089542
VERSION AY089542.1 GI:19528330
KEYWORDS FLI CDNA.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 3934)
Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,B.,
George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,
Miranda,A., Mungall,C.J., Nuncio,J., Pacleb,J., Paragas,V., Park,S.,
Patel,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
and Celnikier,S.
Direct Submission
Submitted (12-MAR-2002) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unsliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our Web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
Location/Qualifiers
1..3934
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/map="88A8-88A9"
1..3934
/gene="foxo"
/note="alignment with genomic scaffold AE003703"
```

```
/db_xref="FLYBASE:FBgn0038197"
785...2626
/gene="foxo"
/note="Longest ORF"
/codon_start=1
/product="LD19191p"
/protein_id="AAL90280.1"
/db_xref="GI:19528331"
/db_xref="FLYBASE:FBgn0038197"
/translation="MMDGYAGEMPRLTHDNGLMDQLGGDLPLDVGFEPPOTRARSNT
WPCRPENFEPTDELSTKASNOQLAPGDSQOAIQNNAKKNSSRRNAGNLSYAD
LI THAIGSATDKRLTLTSQIYEMVMVNYFKDGDSSNSAGKNKSIRHNLGHRFMR
VNEGTSKSWMLNPEAKPKFSVRRRAASMETSYKRGRAKRVLELRQAGVGL
NDATPSSSVSEGLDFHPSLHSGGFGQLSPDFQRASNSASSCCRLSPIRAQDLE
PDMGFFDYONTTMTQAAHALBELTGTMADELTLNQOQOFGAAGSLPQPPPPY
QPPHQAAQQQQQQSPYALNGPASGYNTLQPSQCLLHRSNCSMHNARDGLSPNS
VTTWMSFAYPNSESSLSNTYSNVLDGADTAALVQVQQQQQQQQQLSASLEGQC
LEVLNNEAQPIDEFNLEFPVGNLECNVEBELLQOEMSTGGGLDINIPLATVNTNLVNS
SSGPLSTSNLSNLSNLSNLSNLSNLSNLSNLSNLSNLSNLSNLSNLSNLSNLSN
QQQLLLNNNNSSSSSLELATQTATNLNARVQSPSVTSPSSVWH"
```

ORIGIN

Alignment Scores:

Pred. No.:	2,74e-40	Length:	3934
Score:	450.00	Matches:	78
Percent Similarity:	93.07%	Conservative:	16
Best Local Similarity:	77.23%	Mismatches:	7
Query Match:	81.08%	Indels:	0
DB:	3	Gaps:	0

US-09-844-353A-54 (1-103) x AY089542 (1-3934)

Qy 1 LysLysThrThrThrArgArgAsnAlaTyrGlyAsnMetSerTyrAlaGluLeuThr 20

Db 1040 AAGAAAGAACTCATCGCTCGCAATGCATGCGGAATCTATCTATGCGGATCTCATCG 1099

Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTyrMet 40

Db 1100 CATGCCATTGATCGGCCACGACAAACGATTGACACTGATGATTTACGATGGATG 1159

Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTyrLys 60

Db 1160 GTCCAGATATGCCATATTTCAAGACACAGGCGCATTCGAATAGCAGTCCCGATGGAAG 1219

Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80

Db 1220 AACTCCATACGTCAATCTGCTGTCACACCGCTTTATGAGGGTCCAAACAGAGGC 1279

Qy 81 AlaGlyLysSerSerTyrTyrValIleAsnProAspAlaLysProGlyMetAsnProArg 100

Db 1280 ACCGGCAAGTCATCTGCTGATGCTCAACCCGAGGCGCAAGCGCGCAAGTCTGTGGCG 1339

Qy 101 Arg 101

Db 1340 CGC 1342

RESULT 9

AY092357

LOCUS Homo sapiens cDNA FLJ35038 fis, clone OCEBBF2016841, highly similar to FORKHEAD PROTEIN O3A.

ACCESSION AK092357

VERSION AK092357.1 GI:21750933

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

REFERENCE

AUTHORS Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,

Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isogai,T.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 2685)
Direct Submission
Isogai,T. and Yamamoto,J.
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

TITLE
JOURNAL

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

FEATURES

source

1..2685
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="OCBFF2016841"
/tissue_type="brain"
/clone_lib="OCBFF2"
/dev_stage="fetal"
/note="cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:

Pred. No.: 2,086-39 Length: 2685
Score: 440.50 Matches: 79
Percent Similarity: 90.20% Conservative: 13
Best Local Similarity: 77.45% Mismatches: 9
Query Match: 79.37% Indels: 1
DB: 9 Gaps: 1

US-09-844-353A-54 (1-103) x AK092357 (1-2685)

QY 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuThr 20
Db 146 AGGAATGTCGTCGCGGGGAAACGCTCGGAAACCTGCTCTACCGGACCTGATCACC 205
QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
Db 206 CGCGCCATCGAGAGCTCCCGGCAACACGCTCCTCTGCTCCAGATCTACGAGTGGATG 265
QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db 266 GTGCGTGTGTCCTTACTTCAAGGATAGGGGACAGCAACAGCTCTGCGGCTGGAAG 325
QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db 326 AACTCCATCCGCGACACCTGTCTACTGATAGTCGATTCATGCGGTCAGATGAGGGA 395
QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db 386 ACTGGCAAGAGCTCTTGGTGGATCATCAACCTGTATGGGGGGAAGCGGAAAGCCCC 445
QY 100 ArgArg 101
Db 446 CGGCGG 451
RESULT 10
AK122861
LOCUS
DEFINITION Homo sapiens cDNA FLJ16486 fis, clone BRTHA3007532, moderately
similar to Forkhead box protein O3A.
ACCESSION AK122861
VERSION AK122861.1 GI:34528058
KEYWORDS oligo capping; fis (full insert sequence).

SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

1..2870
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BRTHA3007532"
/tissue_type="thalamus"
/clone_lib="BRTHA3"
/note="cloning vector: pME18SFL3"

FEATURES

source

1..2870
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BRTHA3007532"
/tissue_type="thalamus"
/clone_lib="BRTHA3"
/note="cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:

Pred. No.: 2,246-39 Length: 2870
Score: 440.50 Matches: 79
Percent Similarity: 90.20% Conservative: 13
Best Local Similarity: 77.45% Mismatches: 9
Query Match: 79.37% Indels: 1
DB: 9 Gaps: 1

US-09-844-353A-54 (1-103) x AK122861 (1-2870)

QY 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuThr 20
Db 341 AGGAATGTCGTCGCGGGGAAACGCTCGGAAACCTGCTCTACCGGACCTGATCACC 400
QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
Db 401 CGCGCCATCGAGAGCTCCCGGCAACACGCTCCTCTGCTCCAGATCTACGAGTGGATG 460
QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db 461 GTGCGTGTGTCCTTACTTCAAGGATAGGGGACAGCAACAGCTCTGCGGCTGGAAG 520
QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db 521 AACTCCATCCGCGACACCTGTCTACTGATAGTCGATTCATGCGGTCAGATGAGGGA 580
QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db 581 ACTGGCAAGAGCTCTTGGTGGATCATCAACCTGTATGGGGGGAAGCGGAAAGCCCC 640
QY 100 ArgArg 101
Db 641 CGGCGG 646

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuno,Y., Nagai,K. and Isogai,T.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 2870)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

```
RESULT 11
AF114259      2889 bp      mRNA      linear      ROD 06-JUN-2001
DEFINITION   Mus musculus forkhead protein FKHR2 (Fkhr2) mRNA, complete cds.
ACCESSION    AF114259
VERSION      AF114259.1 GI:5348331
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
              1 (bases 1 to 2889)
REFERENCE    Biggs,W.H. III and Cavenee,W.K.
              Identification and characterization of members of the FKHR (FOX O)
              subclass of winged-helix transcription factors in the mouse
              Mamm. Genome 12 (6), 416-425 (2001)
JOURNAL      21251166
MEDLINE      11353388
REFERENCE    2 (bases 1 to 2889)
              Biggs,W.H. III, Cavenee,W.K. and Arden,K.C.
              Direct Submission
              Submitted (15-DEC-1998) Ludwig Institute for Cancer Research, San
              Diego Branch, 9500 Gilman Drive, La Jolla, CA 92093-0660, USA
JOURNAL
FEATURES
  source
    1..2889
    /organism="Mus musculus"
    /mol_type="mRNA"
    /db_xref="taxon:10090"
    /chromosome="10"
    /dev_stage="E8.5 embryo"
  gene
    1..2889
    /gene="Fkhr2"
  5'UTR
    1..325
    /gene="Fkhr2"
    326..2344
    /gene="Fkhr2"
    /note="member of FKHR-subclass forkhead/winged helix
    family"
  CDS
    /codon_start=1
    /product="forkhead protein FKHR2"
    /protein_id="AAD42107.1"
    /db_xref="GI:5348332"
    /translation="MAEAPASVPPLSPLEVLDPFEFQSPRSCTWPLQRPLOQASP
    AKPSGETAADSMPREDEDEDDGGGRASSAMVIGGVSTLGSGLLEDSAMLLAP
    GQDUGSPASAGALSGTPTLQFPQPQPQAGAGGQPKCSRNRWGNLS
    YADLITRAIESPDRLTLQIYEMWVRCVPYFKDGSNSAGWNSIRHNLHSR
    FMRVQNEGTSKSWIINPDGKSGKAPRRRAVSDNSNKYTKSGRAAKKAAALQAA
    PESADSPQSLSKWPGSPTRSDELDAWTFRSRTNSASTVSGRLSPILASTELDD
    VQDDGPLSPMLYSSASLSPSVKPCTVELPRLTDMAGTNLNDGLAENLMDLDDN
    TALPSPQSPGGLMQRGSSFPYTKASGLSGTGSFNSVTFGSPSLNSLRQSPMOTI
    QENRPATFSSVSHYGNQTLQDLTSLSDLSHSDVMVTQSDPLMSQASTAVSAQNARNV
    MLRNDPMWSFAAQPTQGLVGNQNLHGHQHTQAGLGGSRALNSVSNMGLSSSLGS
    AKHQQSPASQSMQTLSDLSGLSSLYSASANLPVNGHDKFPDLDLDMFNGSLCDME
    SIIRSELMDADGLDFNDSLISTQNVVGLNVGNFTCAKQASSQSWVPG"
  3'UTR
    2347..2889
    /gene="Fkhr2"
ORIGIN
Alignment Scores:
Pred. No.:      2,26e-39      Length:      2889
Score:          440.50      Matches:    79
Percent Similarity: 90.20%      Conservative: 13
Best Local Similarity: 77.45%      Mismatches: 9
Query Match:    79.37%      Indels:    1
DB:             10      Gaps:      1
US-09-844-353A-54 (1-103) x AF114259 (1-2889)
Qy      1 LysLysThrThrThraArgaenAlaTtpGLyAsnMetSerTyrAlaGluLeuIleThr 20
Db      764 AGGAAATGTCCTCGCGGGAATGCTCGGGGAACCTGTCCTATGCCGACCTGATCAC 823
Qy      21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
```

```
Db      824 CGCGCCATCGAGAGCTCCCGGCAAAACGGCTCACTTTGTCCAGATCTACGAGTGGATG 883
Qy      41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db      884 GTGGCTGTGTGGCCCTACTTCAAGGATAAGGGCGACAGCAACAGCTCTGGGGGTGGAAG 943
Qy      61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db      944 AACTTCATCCGCGCAACCTGTCTCTGCACAGCGCTTCATGCGGTTCAAGATGAAGGC 1003
Qy      81 AlaGlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db      1004 ACGGGCAAGAGCTCTTGTGGATCATCAACCCGATGGGGGAAGAGCGGAGGCCCC 1063
Qy      100 ArgArg 101
Db      1064 CGGCGG 1069
RESULT 12
AF032886
LOCUS       AF032886
DEFINITION Homo sapiens forkhead protein (FKHRL1) mRNA, complete cds.
ACCESSION  AF032886
VERSION     AF032886.1 GI:2895493
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 3183)
REFERENCE    Anderson,M.J., Viars,C.S., Czekay,S., Cavenee,W.K. and Arden,K.C.
              Cloning and characterization of three human forkhead genes that
              comprise an FKHR-like gene subfamily
              Genomics 47 (2), 187-199 (1998)
JOURNAL      98140118
MEDLINE      9479491
REFERENCE    2 (bases 1 to 3183)
              Anderson,M.J., Viars,C.S., Czekay,S., Cavenee,W.K. and Arden,K.C.
              Direct Submission
              Submitted (03-NOV-1997) Medicine, Ludwig Institute for Cancer
              Research, San Diego Branch, 9500 Gilman Drive, La Jolla, CA
              92093-0660, USA
JOURNAL
FEATURES
  source
    1..3183
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /chromosome="6"
    /map="6q21"
  gene
    1..3183
    /gene="FKHRL1"
  CDS
    925..2346
    /gene="FKHRL1"
    /codon_start=1
    /product="forkhead protein"
    /protein_id="AAC39592.1"
    /db_xref="GI:2895494"
    /translation="MAEAPASVPPLSPLEVLDPFEFQSPRSCTWPLQRPLOQASP
    AKPSGETAADSMPREDEDEDDGGGRAGSAMVIGGGSGTGLGGLLEDSARVLA
    PGQDPGSGPATAGLGGTQALLQPPLPQPQAGAGGQPKCSRNRWGNL
    SYADLITRAIESPDRLTLQIYEMWVRCVPYFKDGSNSAGWNSIRHNLHS
    RFMRVNEGTSKSWIINPDGKSGKAPRRRAVSDNSNKYTKSGRAAKKAAALQ
    APESADSPQSLSKWPGSPTRSDELDAWTFRSRTNSASTVSGRLSPIMASTELDD
    EVQDDAPLSPMLYSSASLSPSVKPCTVELPRLTDMAGTNLNDGLTENLMDLDD
    NITLPPSPQSPGGLMQRGSSFPYTKASGLSGTGSFNSVTFGSPSLNSLRQSPMOT
    IOENKPATFSSVSHYGNQTLQDLTSLSDLSHSDVMVTQSDPLMSQASTAVSAQNARNV
    MLRNDPMWSFAAQPTQGLVGNQNLHGHQHTQAGLGGSRALNSVSNMGLSSSLGS
    SAKHQQSPVQSQMTLSDLSGLSSLYSTANLPVNGHDKFPDLDLDMFNGSLCDME
    ESIIIRSELMDADGLDFNDSLISTQNVVGLNVGNFTGAKQASSQSWVPG"
ORIGIN
Alignment Scores:
```

Pred. No.:	2,51e-39	Length:	3183
Score:	440.50	Matches:	79
Percent Similarity:	90.20%	Conservative:	13
Best Local Similarity:	77.45%	Mismatches:	9
Query Match:	79.37%	Indels:	1
DB:	9	Gaps:	1
US-09-844-353A-54 (1-103) x AF032886 (1-3183)			
QY	1	LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyAlaGluLeuLeuThr	20
Db	1366	AGGAATGTTCTGTCGCGCGGACGCTGGGAAACCTGCTCTACGCGGACCTGATCACC	1425
QY	21	ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyTGluTrpMet	40
Db	1426	CGGCGCATCAGAGCTCCCGGACAAACGGCTCACTGTGCCAGATCTACGATGGATG	1485
QY	41	ValGlnAsnValProTyPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys	60
Db	1486	GTGCGTGGTGGCCCTACTTCAAGGATAAGGGGACAGACAGCTCTGCGGCTGGAG	1545
QY	61	AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly	80
Db	1546	AATCCATCCGCGACACACCTGTCTCATAGTCGATTCATGCGGGTCCAGATGAGGGA	1605
QY	81	AlaGlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro	99
Db	1606	ACTGGCGAGAGACTCTTGCTGGATCATCAACCTGATGGGGGAGAGCGGAAAGCCCC	1665
QY	100	ArgArg 101	
Db	1666	CGGCGG 1671	
RESULT 13			
BC020227			
LOCUS			
DEFINITION			
Homo sapiens forkhead box O3A, mRNA (cDNA clone MGC:31925			
IMAGE:4111403), complete cds.			
ACCESSION			
BC020227.1 GI:18044232			
VERSION			
MGC.			
KEYWORDS			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Homo sapiens			
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
1 (bases 1 to 3300)			
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,			
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,			
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,			
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,			
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,			
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,			
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,			
Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,			
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,			
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,			
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,			
Villalon,D.K., Munz,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,			
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,			
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,			
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,			
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,			
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalius,D.E.,			
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.			
Generation and initial analysis of more than 15,000 full-length			
human and mouse cDNA sequences			
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
22388257			
PUBMED			
12477932			
2 (bases 1 to 3300)			
Strausberg,R.			
AUTHORS			
Direct Submission			
TITLE			
Submitted (19-DEC-2001) National Institutes of Health, Mammalian			
JOURNAL			
ALIGNMENT SCORES:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
US-09-844-353A-54 (1-103) x BC020227 (1-3300)			
Gene Collection (MGC), Cancer Genomics Office, National Cancer			
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
USA			
NIH-MGC Project URL: http://mgc.nci.nih.gov			
Contact: MGC help desk			
Email: cgapbs-remail.nih.gov			
Tissue Procurement: ATCC			
cDNA Library Preparation: Rubin Laboratory			
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Genome Sequence Centre,			
BC Cancer Agency, Vancouver, BC, Canada			
info@bcsc.bc.ca			
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,			
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,			
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo			
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven			
Ness, Pawan Pandoh, Anna-Liisa Prabhur, Parvaneh Saedi, Jacqueline			
Schein, Duane Smalus, Michael Smith, Lorraine Spence, Jeff Scott,			
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,			
George Yang, Scott Zuyderduyn, Marco Marra.			
Clone distribution: MGC clone distribution information can be found			
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov			
Series: IRAL Plate: 40 Row: c Column: 5			
This clone was selected for full length sequencing because it			
passed the following selection criteria: matched mRNA gi: 4503738.			
FEATURES			
Location/Qualifiers			
1..3300			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="MGC:31925 IMAGE:4111403"			
/tissue_type="Muscle, rhabdomyosarcoma"			
/clone_lib="NIH MGC 17"			
/lab_host="DH10B-R"			
/note="Vector: POTB7"			
1..3300			
/genes="FOXO3A"			
/notes="synonyms: FKHL1, AF6q21, FKHLIP2"			
/db_xref="LocusID:2309"			
/db_xref="MIM:602681"			
305..2326			
/codon_start=1			
/product="FOXO3A protein"			
/protein_id="AAH20227.1"			
/db_xref="GI:18044233"			
/db_xref="LocusID:2309"			
/translation="MAEAPASAPLSPLEVLDPPEPQSPRSCWPIQRLQASP			
APSGETAADSMIPEDDEDEDDGGAGSMAIGGGGGSGTLGGLLLEDSARVLA			
PGQDPGSGFATAAGLGGTQALLQPPQPPQGAAGSGGQPKKCSRRNAGNL			
SYADLITRAIESDPKRLTSLQIYEMVRCVPYFKDKGDSNAGWKNIRHLSLHS			
RPMRVQNEGTGKSSWIIINPDGKGKAPRRASVMDNSNKTYSRGAARKKALQT			
APESADDSPLSKWPGSPTRSSDELDAWTFRSTRNSASTVSGRLSPIMASTELD			
EVQDDAPLSPLMYSSASLSVSKPCTVELPLRTDMAGTMNDGLTENLMDLLD			
NITLPPSPSPFTGLMQRSSSPYTKSGSLGSPSSFSFSTVFGPSLMSKQSPMQT			
IQENKPAFTSSMHYQNTQLDQLTSDLSHSDVMMVMTQSDPLMSQASTAVASQSRN			
VMRNDPMMSFAAQNGSLVNLHGHQHTQAGLGGSRALSNVSNMGLSESSILG			
SAKHQQSPVSMQTLSDLSGSLYSTANLPVNGHEKFFPSDLDDMFNGSLSCDM			
ESIRIIRSELMADGLDFNFDLSLSTQNVLVGNFTGAQASQSSQSVPG"			
767..1021			
/note="FH; Region: FORKHEAD"			
/db_xref="CDD:smart00339"			
misc_feature			
2,62e-39			
Length:			
Matches:			
79			
Conservative:			
13			
Mismatches:			
9			
Indels:			
1			
Gaps:			
9			

Qy 1 LysLysThrThrArgArgAsnAlaTrrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
 Db 746 AGGAAATGTTCTGTCGGCGGGAACGCTGGGGAAACCTGTCTTACGGGACCTGATCACC 805
 Qy 21 ThrAlaIleMetAlaSerProGluIlyAargLeuThrIleuAlaGlnValTyrGluTrrpMet 40
 Db 806 CGCGCCATTCGAGAGCTCCCGGCAACACGGCTCACTCTGTCCAGATCTACAGTGGATG 865
 Qy 41 ValGlnAsnValProTrrpPheAargAspIlyGlyAspSerAsnSerSerAlaGlyTrrpIlys 60
 Db 866 GTGCGTTCGTCGCTCACTTCAAGGATAAGGCGACCAACAGCTCTGCGCGCTGGAAG 925
 Qy 61 AsnSerIleAargHisAsnLeuSerIleuHisSerAargPheMetAargIleGlnAsnGluCly 80
 Db 926 AACTCCATCCGCGCAACCTGTCACTGTCATAGTGAATTCATGCGGGTCCAGAAAGAGGA 985
 Qy 81 AlaGlyLysSerSerTrrpTrrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
 Db 986 ACTGGCAAGAGCTTCTGTGATCATCAACCTGATGGGGAGAGCGGAAAGCCCC 1045
 Qy 100 ArgArg 101
 Db 1046 CGGCGG 1051
 RESULT 14
 LOCUS BC021224
 DEFINITION Homo sapiens forkhead box O3A, mRNA (CDNA clone MGC:12739
 IMAGE:4137370), complete cds.
 ACCESSION BC021224
 VERSION BC021224.2 GI:33871764
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 3300)
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heideh,F.,
 Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Tohiyuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Iqbal,M., N.A., Peters,G.J.,
 Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whitting,M., Madan,M., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,
 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 MEDLINE 12477932
 PUBLISHED
 REFERENCE 2 (bases 1 to 3300)
 Strausberg,R.
 AUTHORS
 TITLE Direct Submission
 JOURNAL Submitted (14-JAN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 On Aug 19, 2003 this sequence version replaced gi:18203784.
 COMMENT Contact: MGC help desk
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mc@nhgri.nih.gov
 Akhter,N., Ayale,K., Beckettrom-Sternberg,S.M., Benjamin,B.,
 Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghghi,P.,
 Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
 Maduro,Q.L., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
 McDowell,J., Pearson,C., Stanttripol,S., Thomas,P.J., Touchman,J.W.,
 Tsurgon,C., Voigt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
 Young,A., Zhang,L.-H. and Green,E.D.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 17 Row: j Column: 2
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4503738.

FEATURES

source

1..3300

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:12739 IMAGE:4137370"

/tissue_type="Muscle, rhabdomyosarcoma"

/clone_lib="NIH MGC 17"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

1..3300

/gene="FOXO3A"

/note="synonyms: FKHL1, AF6q21, FKHL1P2"

/db_xref="LocusID:2309"

/db_xref="MIM:602681"

305..2326

/codon_start=1

/product="FOXO3A protein"

/protein_id="AAH21224.1"

/db_xref="GI:18203785"

/db_xref="LocusID:2309"

/translation="MAEAPASPAFLSPLEVLDPFEFQSPRSTWLPQPELQASP
 AKPSGTAADSMIPEDDEDDGGRAGSAMAIGGGSGTGTGSGLLLEDSDARVLA
 PGGDDPGSPATAAGLGGTQALLOPPLPPQPPQGAAGSGGCPKCKSSRNAGNL
 SYALDLTRALESSPDKLTLSQIYEMVRCVYFVKDGSNSAGWKNISRHLSLHS
 RFMRVQNEGTSKSWIINPDGKSGKAPRRAYMDNSNXYTKSRGAAKKALQT
 APESADDSPLSKWPGSPERSDELDAMTDFRSRNSNASTSVSGRLSPIMASTELL
 EVQDDAPLSPMLYSSASLSPSVSKPTVELPRLTDMAGTMNLDGLTENLMDLLD
 NITLPPSPGPTGGLMQRSSSPFYTKGSLGSPSTSFNSFVFGPSLSLRQSPMOT
 IOENKPTFSSMSHYQTLQDLTSDLSHSDVMTQSDPLMSQASTAVSAQNSRRN
 VMLRNDPMSPFAOPNOSLQNLHHQHOTGALGGSRLSNSVSNMGLSESSSLG
 SAKHQQSPVSQMTLSDSLSSGSLSYSTANLPMVHGKPTPSLDLDMFNGSLBCDM
 ESIIRSELMDADGLDFNFDLSLSTQNVVGLNVGNFTGAKQASSQSWVEG"

767..1021

/note="FH; Region: FORKHEAD"

/db_xref="CDD:smart00339"

misc_feature

767..1021

/note="FH; Region: FORKHEAD"

/db_xref="CDD:smart00339"

ORIGIN

Alignment Scores:

Pred. No.:

2.62e-39 Length: 3300

Score: 440.50 Matches: 79

Percent Similarity: 90.20% Conservative: 13

Best Local Similarity: 77.45% Mismatches: 9

Query Match: 79.37% Indels: 1

DB: 9 Gaps: 1

US-09-844-353A-54 (1-103) x BC021224 (1-3300)

Qy 1 LysLysThrThrArgArgAsnAlaTrrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
 Db 746 AGGAAATGTTCTGTCGGCGGGAACGCTGGGGAAACCTGTCTTACGGGACCTGATCACC 805

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 29, 2004, 09:24:12 ; Search time 341 Seconds
(without alignments)
1283.180 Million cell updates/sec

Title: US-09-844-353A-54
Perfect score: 555
Sequence: 1 KKTTRRNAGNWSYAELIT.....SSWWINPDAPGCMNPRTR 103

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 segs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xl
-Q=/cgn2_1/USPTO.spool_p/US09844353/runat_23072004_084913_28228/app_query.fasta_1.263
-DB=N Geneseq_29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09844353 @CGN_1_885 @runat_23072004_084913_28228 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_29Jan04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002s.*
7: Geneseqn2003as.*
8: Geneseqn2003bs.*
9: Geneseqn2003cs.*
10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	555	100.0	3499	3	AAG62298 Caenorhab
2	427.5	77.0	5723	6	ABV94143 Breast ca
3	422.5	76.1	2067	7	ACD19349 cDNA enco
4	410.5	74.0	1073	4	AAS00244 Fusion pr
5	410.5	74.0	3387	4	AAI60219 Human pol
6	410.5	74.0	3394	4	AAI58433 Human pol
7	410.5	74.0	3394	8	ADB48404 Novel hum
8	406.5	73.2	3171	6	ABS73211 DNA encod

9	406.5	73.2	3171	6	ABS73212	AbS73212 DNA encod
10	385.5	69.5	5833	5	AAS85053	Aas85053 DNA encod
11	378.5	68.2	2704	3	AAA62299	Aaa62299 Caenorhab
12	335.5	60.5	1353	7	ABZ58161	Abz58161 Human tra
13	335	60.4	2162	4	ABL05685	AbL05685 Drosophil
14	262	47.2	25891	4	ABL05684	AbL05684 Drosophil
15	222.5	40.1	5123	7	AAD55840	Aad55840 Human nuc
16	195	35.1	2404	4	ABL19649	AbL19649 Drosophil
17	195	35.1	2423	4	ABL19651	AbL19651 Drosophil
18	195	35.1	2529	4	ABL18187	AbL18187 Drosophil
19	189.5	34.1	834	4	AAS26564	Aas26564 Human cdn
20	189.5	34.1	834	4	AAS26146	Aas26146 Human cdn
21	189.5	34.1	834	5	ABAI1695	AbAI1695 Human nov
22	189.5	34.1	834	7	ABX73487	Abx73487 Human nov
23	189.5	34.1	834	7	ABX73905	Abx73905 Human nov
24	188.5	34.0	2448	9	ADD29815	Add29815 Human tum
25	188.5	34.0	2559	9	ADD15216	Add15216 Human ser
26	188.5	34.0	2832	2	AAQ50631	Aaq50631 Brain fac
27	188.5	34.0	2934	6	ABI99306	Abi99306 Mouse isc
28	187.5	33.8	1098	4	ABL02111	AbL02111 Drosophil
c	186	33.5	1902	4	ABL05301	AbL05301 Drosophil
	186	33.5	3902	4	ABL05300	AbL05300 Drosophil
	184.5	33.2	1860	2	AAQ50632	Aaq50632 Brain fac
31	184.5	33.1	2299	3	AAF16154	Aaf16154 Human pro
32	183.5	32.9	1662	2	AAZ31672	Aaz31672 Human for
33	182.5	32.9	1662	2	AAZ38080	Aaz38080 Human for
34	182.5	32.9	1800	4	ABL17373	AbL17373 Drosophil
35	182.5	32.9	1976	2	AAZ28095	Aaz28095 Human FRE
36	182.5	32.9	2271	6	ABZ35021	Abz35021 Human gen
37	182.5	32.9	2487	9	ADB75311	Adb75311 Prostata
38	182.5	32.9	3800	4	ABLI17372	AbLI17372 Drosophil
39	182.5	32.9	3946	2	AAZ31671	Aaz31671 Human FKH
40	182.5	32.9	3946	2	AAZ38079	Aaz38079 Human for
41	182.5	32.9	5181	6	ABLI68935	AbLI68935 Kidney ca
42	182.5	32.9	2737	7	AAI51229	Aai51229 Human Fox
43	182	32.8	3342	2	AAV37494	Aav37494 Human hep
44	182	32.8	3397	7	ABX76161	Abx76161 Lung canc
45	182	32.8				

ALIGNMENTS

RESULT 1
AAA62298
ID AAA62298 standard; cDNA; 3499 BP.
XX
AC
XX
AC
XX
DT 11-JAN-2001 (first entry)
XX
DE Caenorhabditis elegans daf-16 cDNA #1.
XX
KW Caenorhabditis elegans; daf-16; daf-2; age-1; daf-18;
KW insulin signalling pathway; insulin receptor; PI 3-kinase; PKB kinase;
KW AKT kinase; PTEN lipid phosphatase; antidiabetic; anorectic; obesity;
KW diabetes; impaired glucose tolerance; transgenic animal; ss.
XX
OS Caenorhabditis elegans.
XX
FH
FH
FT Key Location/Qualifiers
FT CDS 325..1857
FT /*tag= a
FT /product= "DAF-16"
FT /transl_except= (pos:679..681,aa:Pro)
FT /transl_except= (pos:682..684,aa:Ser)
FT /transl_except= (pos:685..687,aa:Asp)
FT /transl_except= (pos:1207..1208,aa:Asp)
FT /note= "the codon at position 1207 to 1208 has an
apparent 1 nucleotide deletion, which alters the reading
frame"

WO200033068-A1.

08-JUN-2000.


```
Db      836 AAGCAGAGCTGTCCTCCGCGCAACGGCTGGGGCAACTGCTTACCGCGCACTCATCAC 895
QY      21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
Db      896 AAGGCCATCGAGAGCTCGGCGGAGAGCGGCTCACGCTGCGCAGATCTACGAGTGGATG 955
QY      41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerAlaGlyTrpLys 60
Db      956 GTCAGAGCGTGCCTCTCTTCAAGGATAGGGTGACAGCAACAGCTCGGCGGGCTGGAAG 1015
QY      61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db      1016 AATTCATTCGTCATATCTGTCCTACACAGCAAGTTCATTCGTGTGCGAGATGAGGA 1075
QY      81 AlaGlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db      1076 ACTGGAAAAAGTTCTTGGTGGATGCTCAATCCAGAGGGTGGCAAGACGGGAAATCTCCT 1135
QY      100 ArgArg 101
Db      1136 AGGAGA 1141

RESULT 3
ID      ACD19349
XX      ACD19349 standard; cDNA; 2067 BP.
AC      ACD19349;
XX      ACD19349;
DT      25-AUG-2003 (first entry)
DE      cDNA encoding novel human protein #29.
XX      Human; NOV; gene therapy; endocrine related disease; diabetes;
KW      metabolism-related disease; obesity; central nervous system disorder;
KW      Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;
KW      schizophrenia; depression; autoimmune disorder; inflammatory disorder;
KW      psoriasis; allergy; lupus erythematosus; asthma; cancer;
KW      inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;
KW      colon cancer; lung cancer; liver cancer; breast cancer; ovarian cancer;
KW      prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;
KW      lung disease; emphysema; obstructive pulmonary disease; haemophilia;
KW      stroke; infection; gene; ss.
XX      Homo sapiens.
OS      Homo sapiens.
XX      WO2003023002-A2.
XX      20-MAR-2003.
XX      09-SEP-2002; 2002WO-US028539.
XX      07-SEP-2001; 2001US-0318120P.
XX      07-SEP-2001; 2001US-0318130P.
XX      10-SEP-2001; 2001US-0318430P.
XX      17-SEP-2001; 2001US-0322636P.
XX      17-SEP-2001; 2001US-0322781P.
XX      17-SEP-2001; 2001US-0322816P.
XX      17-SEP-2001; 2001US-0322817P.
XX      19-SEP-2001; 2001US-0323519P.
XX      20-SEP-2001; 2001US-0323631P.
XX      20-SEP-2001; 2001US-0323636P.
XX      25-SEP-2001; 2001US-0324969P.
XX      25-SEP-2001; 2001US-0325091P.
XX      26-SEP-2001; 2001US-0324990P.
XX      17-APR-2002; 2002US-0373212P.
XX      06-SEP-2002; 2002US-00236171.
XX      (CURA-) CURAGEN CORP.
PA      Spytek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;
PI      Gerlach VL, Vernet CAM, Ellerman K, Berghs C, Rothenberg ME, Guo X;
PI      Shimkets RA, Leach WD, Catterton E, Kekuda R, Ji W, Miller CE;
PI      Rieger DK, Taupier RJ, Shenoy SG, Liu X, Padigaru M, Alsbrook JP;
```

```
PI      Lepley DM, Edinger SR, Burgess CE;
XX      WPI; 2003-313242/30.
DR      P-PSDB; ABO14656.
XX      New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOVX)
PT      and polynucleotides, useful in gene therapy, e.g. for treating or
PT      preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,
PT      stroke or infections.
XX      Claim 20; Page 145; 586pp; English.
XX      The invention describes a new isolated polypeptide (NOVX). The NOVX
CC      polypeptide, nucleic acid and antibody are useful as therapeutics,
CC      particularly in the manufacture of a medicament for treating a syndrome
CC      associated with a human disease, which includes a pathology associated
CC      with NOVX polypeptide. The DNA encoding the protein is useful in gene
CC      therapy for treating the disease or condition. In particular, the NOVX
CC      polypeptide or polynucleotide is useful for treating endocrine/
CC      metabolism-related diseases (e.g. obesity or diabetes), central nervous
CC      system disorders (e.g. Alzheimer's disease, Parkinson's disease,
CC      epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune
CC      and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,
CC      asthma, inflammatory bowel disease, rheumatoid arthritis or
CC      osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,
CC      prostate or brain cancers, or melanoma), liver diseases (e.g. liver
CC      cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),
CC      haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
CC      These are also useful in developing powerful assay system for functional
CC      analysis of various human disorders, as well as in diagnostic
CC      applications, and for monitoring the effects of drugs during clinical
CC      trials. This sequence encodes a novel human NOV protein
XX      SQ      Sequence 2067 BP; 480 A; 642 C; 553 G; 391 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.:      1,67e-45      Length:      2067
Score:          422.50      Matches:      76
Percent Similarity: 90.53%      Conservative: 10
Best Local Similarity: 80.00%      Mismatches: 8
Query Match:      76.13%      Indels:      1
DB:              7      Gaps:      1

US-09-844-353A-54 (1-103) x ACD19349 (1-2067)
QY      8 AsnAlaTrpGlyAsnMetSerTyrAlaGluLeuThrAlaIleMetAlaSerPro 27
Db      271 AACGCTGGGAAACCTGCTCTACGGGACCTGATCACCGCGCATCCGAGCTCCCCG 330
QY      28 GluLysArgLeuThrLeuAlaGlnValTyrGluTrpMetValGlnAsnValProTyrPhe 47
Db      331 GACAAACGGCTCACTCTGTCCAGATCTACGAGTGGATGGTGGCTGCTGCTACTTC 390
QY      48 ArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLysAsnSerIleArgHisLeu 67
Db      391 AAGGATAGGGCGAGCAGCAACAGCTCTCGCGCTGGAGAACTCCATCCGGCACACCTG 450
QY      68 SerLeuHisSerArgPheMetArgIleGlnAsnGluGlyValaGlyLysSerTrpTrp 87
Db      451 TCACCTGCATATCGATTTCATCGGGGTCCAGATGAGGAACTGGCAGAGCTCTTGGTGG 510
QY      88 ValIleAsnProAsp---AlaLysProGlyMetAsnProArgArg 101
Db      511 ATCATCAACCTGTATGGGGGGAAGAGCGGAAAGCCCCCGCGG 555

RESULT 4
AAS00244
ID      AAS00244 standard; DNA; 1073 BP.
XX      AAS00244;
XX      AAS00244;
DT      09-MAY-2001 (first entry)
XX      XX
```

Fusion protein GST-AfxDBD, DNA sequence.

DE XX DNA binding domain; Afx; transcription factor; GST-AfxDBD; ds;
KW human fork head transcription factor; Afx response element; human;
XX diabetes; drug target; insulin; insulin receptor signalling pathway.
KW XX Homo sapiens.
OS Synthetic.
OS Chimeric.

Key Location/Qualifiers
CDS 1..1073
FT /*tag= a
FT /product= "GST-AfxDBD fusion protein"
FT /transl_except= (pos:82..84, aa:Thr)
FT /transl_except= (pos:1046..1047, aa:Ser)
FT /note= "This codon contains an apparent 1 nucleotide
FT deletion which alters the reading frame"
XX XX

WO200114544-A1.
XX 01-MAR-2001.
XX 22-AUG-2000; 2000WO-SE001603.
XX 26-AUG-1999; 99SE-00003009.
XX 31-AUG-1999; 99US-0151867P.
XX 25-NOV-1999; 99SE-00004269.
XX (PHAA) PHARMACIA & UPJOHN AB.
XX Clment-Johansson I, Dahlman-Wright K, Lake S, Wasserman W;
XX WPI; 2001-218446/22.
XX P-PSDB; AAU00214.
XX New Afx response element with a nucleotide sequence comprising a DNA
XX binding site for the human fork head transcription factor Afx, useful in
XX screening for genes or in bioinformatic analysis of the human genome.
XX Example 2; Fig 7; 34pp; English.

The sequence represents the coding sequence of the GST-AfxDBD fusion
protein encoded by expression vector pGEX-AfxDBD. Human fork head
transcription factor, Afx, was expressed and the protein used to find a
response element comprising an 8 base pair (bp) nucleotide sequence. The
nucleotide sequence comprises AACATGTT, the Afx response element, and is
useful in bioinformatic analysis e.g. of the human genome. Employing the
Afx response element is also useful for screening for genes that may be
used as diabetes drug targets. This can provide a subset of genes
transcriptionally responsive to insulin and may lead to development of
assays that facilitate the analysis of genes interacting with the insulin
receptor pathway. The genes found in such screening may also be employed
in further screening methods for compounds which modify the insulin
receptor signalling pathway

SQ Sequence 1073 BP; 300 A; 235 C; 268 G; 270 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.56e-44 Length: 1073
Score: 410.50 Matches: 73
Percent Similarity: 86.27% Conservative: 15
Best Local Similarity: 71.57% Mismatches: 13
Query Match: 73.96% Indels: 1
DB: 4 Gaps: 1

US-09-844-353A-54 (1-103) x AAS00244 (1-1073)

Qy 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
Db 715 CGGAAGGAGGAGGCTCCCGCGGAATGCTGGGAAATCAGTCTATGCGAGACTATCAGC 774
Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40

Db 775 CAGGCCATTGAAAGCGCCCGGAGAGCGGACTGACACTTCCCCAGATCTACGAGTGATG 834
Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db 835 GTCGCTACTTACCTTCAAGGACAAAGGGTGACAGCAACAGCTCAGCAGGATGGAAG 894
Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db 895 AACTCGATCCGCCAACACCTGTCCCTGCAGCAAGTTTCATCAGGTTTCACACGAGGCC 954
Qy 81 AlaGlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db 955 ACCGGCAAAAGCTCTTGGTGAGTCTGACCCCTGAGGAGGCAAGAGCGGCAAGCCCC 1014
Qy 100 ArgArg 101
Db 1015 CGCGGC 1020

RESULT 5
AAI60219
ID AAI60219 standard; cDNA; 3387 BP.
XX AC
XX AAI60219;
XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4208.
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX Homo sapiens.
XX WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US034263.
XX 23-DEC-1999; 99US-00471275.
XX 21-JAN-2000; 2000US-00488725.
XX 25-APR-2000; 2000US-00552317.
XX 20-JUN-2000; 2000US-00598042.
XX 19-JUL-2000; 2000US-00620312.
XX 03-AUG-2000; 2000US-00653450.
XX 14-SEP-2000; 2000US-00662191.
XX 19-OCT-2000; 2000US-00693036.
XX 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang Y, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX P-PSDB; AAM41063.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX Claim 1; SEQ ID NO 4208; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous

Db 990 CGCGC 995

RESULT 7

ID ADB48404 standard; cDNA; 3394 BP.

XX ADB48404;

AC ADB48404;

XX 04-DEC-2003 (first entry)

DT Novel human cDNA SEQ ID NO 314.

DE ss; cancer; neurodegenerative disease; human.

KW Homo sapiens.

OS US2003104529-A1.

XX 05-JUN-2003.

XX 04-JAN-2002; 2002US-00037270.

XX 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR 19-JUL-2000; 2000US-00620312.

XX (ZHOU/) ZHOU P.

PA (TANG/) TANG Y T.

PA (LIUC/) LIU C.

PA (ASUN/) ASUNDI V.

PA (DRMA/) DRMANAC R T.

XX Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;

PI WPI; 2003-678194/64.

DR New polynucleotide, useful for treating diseases e.g., cancer or neurodegenerative diseases.

PT Claim 1; SEQ ID NO 314; 99pp; English.

XX The invention relates to a polynucleotide comprising a sequence given in the specification, or its mature protein-coding portion, or its complement. The polynucleotide is useful for treating diseases e.g., cancer or neurodegenerative diseases and many others listed in the specification. The present sequence represents a novel human cDNA. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030104529.

CC Seq 3394 BP; 790 A; 893 C; 999 G; 712 T; 0 U; 0 Other;

SQ

Alignment Scores:

Pred. No.:	1.32e-43	Length:	3394
Score:	410.50	Matches:	73
Percent Similarity:	86.27%	Conservative:	15
Best Local Similarity:	71.57%	Mismatches:	13
Query Match:	73.96%	Indels:	1
DB:	8	Gaps:	1

US-09-844-353A-54 (1-103) x ADB48404 (1-3394)

Qy 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAenMetSerTyrAlaGluLeuIleThr 20

Db 690 CGGAAGGAGGCTCCGCGGAGATGCTGGGGAANTCACTATATGAGACTATCAGC 749

Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40

Db 750 CAGGCCATTGAAGGCGCCGAGAGGACCTGACACCTGCCAGATCTACGAGTGGATG 809

Qy 41 ValGlnAnValProTyrPheArgAspLysGlyAspSerAnSerSerAlaGlyTrpLys 60

Db 810 GTCCGTACTGTACTTCAAGGACAGAGGGTGACAGCAACAGCTCAGAGGATGGAG 869

Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80

Db 870 AACTCGATCCGCACACACCTGTCCCTGCACAGCAAGTTTCATCAAGTTTCACACGAGCC 929

Qy 81 AlaGlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99

Db 930 ACCGGCAAAAGCTCTTGGTGATGCTGACCTGAGGAGGCAAGAGCGCAAGAGCCCC 989

Qy 100 ArgArg 101

Db 990 CGCGC 995

RESULT 8

ID ABS73211 standard; DNA; 3171 BP.

XX ABS73211;

XX 04-DEC-2002 (first entry)

DT DNA encoding human translocation (X; 11)(ql3; q23) protein #1.

DE Chromosome aberration; oncogenic fusion protein; cancer; proliferative disease; cellular protein isoform; heat shock protein 90; Hsp-90; rheumatoid arthritis; cancer; haematopoietic disorder; T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML; acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML; acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour; papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma; rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.

OS Homo sapiens.

XX WO200269900-A2.

XX 12-SEP-2002.

XX 01-MAR-2002; 2002WO-US006518.

XX 01-MAR-2001; 2001US-0272751P.

XX (CONF-) CONFORMA THERAPEUTICS CORP.

XX Fritz LC, Burrows FJ;

XX WPI; 2002-698710/75.

XX P-PSDB; ABG95042.

PT Treating genetically-defined disease associated with chromosomal aberrations yielding oncogenic fusion proteins, e.g. cell proliferative diseases, involves administering an inhibitor of heat shock protein 90.

PT Disclosure; Page 154-155; 389pp; English.

XX The invention describes a method of treating genetically-defined disease associated with chromosomal aberrations yielding oncogenic fusion proteins (I), treating cancerous cells containing (I) in a heterogeneous cell population, treating proliferative diseases associated with mutant protein or cellular protein isoforms (II) dependent on heat shock protein (HSP)-90, or selectively treating cells expressing (II) involving administering HSP90-inhibitor. The method is useful for treating genetically-defined disease with chromosomal aberration yielding oncogenic fusion protein, treating cancerous cells containing fusion protein in heterogeneous cell population, treating proliferative disease (e.g. rheumatoid arthritis or cancer) associated with mutant protein or cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g. p53), or selectively treating cells expressing mutant protein or cellular protein isoform in a patient heterozygous for (II). The method is useful for treating a disease e.g. haematopoietic disorder such as T or B cell lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CMML, or a disease characterised by a solid tumour such as papillary thyroid carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and


```
XX SQ Sequence 2704 BP; 709 A; 669 C; 452 G; 874 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1.68e-39 Length: 2704
Score: 378.50 Matches: 67
Percent Similarity: 75.93% Conservative: 15
Best Local Similarity: 62.04% Mismatches: 21
Query Match: 68.20% Indels: 5
DB: 3 Gaps: 1

US-09-844-353A-54 (1-103) x AAA62299 (1-2704)
QY 1 LysLysThrThr-----ThrArgAsnAlaTrpGlyAsnMetSerTyr 15
Db 398 AAGNAGCAACCGATCAATTGGCGCAGAGAACCAGATCCATGGGTGAGGAATCCCTAT 457
QY 16 AlaGluLeuThrThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGln 35
Db 458 TCGGATATCATTTGCCAAGCATTTGGATCGCGCCAGCAGAGGCTTAACTCAATGAG 517
QY 36 ValTyrGluTrpMetValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSer 55
Db 518 ATTATCAATGGTTCTCTGATAATATTCCCTACTTTGGAGACGATCTAGTCCCGAGGAG 577
QY 56 SerAlaGlyTrpLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArg 75
Db 578 GCCGCGGATGGAGAACTCGATCCGTCACAACTGTCCTTCAATCTCGTTTCGCGA 637
QY 76 IleGlnAsnGluGlyAlaGlyLysSerSerTrpTrpValIleAsnProAspAlaLysPro 95
Db 638 ATTCAGAAATGAAGAGCGGAAAGAGCTCGTGGTGGTTATTATCCAGATGCAGAGCCA 697
QY 96 GlyMetAsnProArgArgThrArg 103
Db 698 GGAATCAATCCACGGCGTACACGT 721

RESULT 12
ABZ58161
ID ABZ58161 standard; cDNA; 1353 BP.
XX AC ABZ58161;
XX DT
XX 22-APR-2003 (first entry)
XX DE Human transcription factor AFX-zeta splice variant cDNA.
XX KW AFX-zeta; human; transcription factor; splice variant; antidiabetic;
XX KW antiarteriosclerotic; antilipaeamic; cytostatic; antiarthritic; nootropic;
XX KW neuroprotective; anticonvulsant; antiparkinsonian; anorectic;
XX KW hypotensive; nephrotropic; cardiant; gene therapy; gene; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 1..1353
XX FT /*tag= a
XX ET /product= "Human AFX-zeta"
XX FN
XX WO2003002601-A2.
XX PD
XX 09-JAN-2003.
XX PF 28-JUN-2002; 2002WO-BP007160.
XX PR 29-JUN-2001; 2001US-0302134P.
XX PA (NOVS ) NOVARTIS AG.
XX PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX PI Bowen BR, Yang Z, Whelan JP;
XX DR WPI; 2003-210242/20.
```

```
XX PT New human AFX-zeta polypeptide, useful for identifying substances that
PT antagonize or agonize the activity of AFX-zeta protein for treating a
PT disease associated with AFX-zeta protein, e.g. diabetes, atherosclerosis
PT or dyslipidemia.
XX PS Claim 5; Page 35; 48pp; English.
XX CC The present sequence is the coding sequence for AFX-zeta, a novel splice
CC variant of human transcription factor AFX. AFX-zeta was cloned from human
CC liver and heart libraries. The encoded protein has a deletion of 55 amino
CC acids as compared with AFX-alpha protein (see ABP72186), including the
CC first 16 amino acids of the forkhead domain of the AFX-alpha protein.
CC Reporter gene assays showed that AFX-zeta is a potent transcription
CC activator with properties distinct from AFX-alpha. It is regulated by the
CC insulin signalling pathway, binding insulin responsive sequence (IRS)
CC fragments, and by an agent known to affect AMPK activity. It may be a
CC distal effector of the insulin-signalling pathway. AFX-zeta mRNA shows
CC strongest expression in liver, kidney and pancreas. The invention
CC provides assay techniques for identifying human AFX-zeta agonists and
CC antagonists. These can be used to treat and/or prevent insulin
CC resistance, hyperglycaemia, hypoglycaemia, hepatic gluconeogenesis,
CC hypercholesterolaemia, obesity, type 2 diabetes, impaired glucose
CC tolerance, disorders related to syndrome X including hypertension,
CC obesity, insulin resistance, coronary artery disease, glomerulonephritis,
CC glomerulosclerosis, nephrotic syndrome and hypertensive nephrosclerosis,
CC cancer, cancer metastasis, arthritis, parkinson's disease, Alzheimer's
CC disease, Huntington's disease, atherosclerosis or dyslipidaemia
XX SQ Sequence 1353 BP; 295 A; 445 C; 355 G; 258 T; 0 U; 0 Other;
```

Alignment Scores:

```
Pred. No.: 3.19e-34 Length: 1353
Score: 335.50 Matches: 59
Percent Similarity: 87.65% Conservative: 12
Best Local Similarity: 72.84% Mismatches: 9
Query Match: 60.45% Indels: 1
DB: 7 Gaps: 1
```

US-09-844-353A-54 (1-103) x ABZ58161 (1-1353)

```
QY 22 AlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMetVal 41
Db 172 GCCATTGAAGAGCGCCCGGAGAGCGACTGCACACTGCCAGATTTACGATGGATGTC 231
QY 42 GlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLysAsn 61
Db 232 CGTACTGTACCTACTTCAAGACCAAGGTGACGACCAACAGCTCAGCAGATGGAGAAC 291
QY 62 SerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGlyAla 81
Db 292 TCGATCCGCGCACAACTGTCTCCCTGCACAGCAAGTTTCATCAAGGTTTCACAAACGAGCCACC 351
QY 82 GlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnProArg 100
Db 352 GGCAGAAAGCTCTTGGTGGATGCTGAACCTGTAGGAGCGACAGAGCGGCAAGCCCCCGC 411
QY 101 Arg 101
Db 412 CGC 414
```

RESULT 13

```
ABL05685
ID ABL05685 standard; cDNA; 2162 BP.
XX AC ABL05685;
```

XX AC ABL05685;

XX DT 26-MAR-2002 (first entry)

```
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11537.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
```

XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US009231.
 XX PR 23-MAR-2000; 2000US-0191637P.
 XX PR 11-JUL-2000; 2000US-00614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX DR P-PSDB; ABB61582.
 XX DR New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX PS Claim 1; SEQ ID NO 11537; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
 CC AB572072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 2162 BP; 574 A; 602 C; 553 G; 433 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 7,24e-34 Length: 2162
 Score: 335.00 Matches: 78
 Percent Similarity: 33.57% Conservative: 16
 Best Local Similarity: 27.86% Mismatches: 7
 Query Match: 60.36% Indels: 179
 DB: 4 Gaps: 1
 US-09-844-353A-54 (1-103) x ABL05685 (1-2162)
 QY 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
 DB 778 AAGAGAACTATCGCTGCGCAATGATGCGGGAATCTATCTATGCGGATCTCATCAG 837
 QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
 DB 838 CATGCCATGGATGGCCGCCACGACAAAGCAATGACACTGAGTCAGATTACAGTGGATG 897
 QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
 DB 898 GTCCAGAATGTCCATATTTCAAGGACAGGCGGATTCGATAGCAGTCCCGAATGGAG 957
 QY 60 ----- 60
 DB 958 AGTTTTCGAATTAATGCGCTTTTATCTTCGGATCGTTTTCGTTTGTGTTTACATCGCAGC 1017
 QY 60 ----- 60
 DB 1018 ATACCACCTCTGGCTGTGATAAAGACCAACGCTGTTGAAGAAATCGCTGCGCTA 1077
 QY 60 ----- 60
 DB 1078 CGAAGAAATCGAAGACGACTCTTCCACTTGATACCGAAGAACTTATTTTCGACAGCAGC 1137
 QY 60 ----- 60

Db 1138 GCGTTTCCAGCAATTTTCAGCTTCTTCTCCGCGGGTTGGTGGTTCGGTGGTTCGGT 1197
 QY 60 ----- 60
 Db 1198 GGTTCGGGATCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1257
 QY 60 ----- 60
 Db 1258 TCCTCGCAGTCGGGCGGATTTGGCCATTGAACCGGTGCCAAAATTTTACGAGTTGAGCAGC 1317
 QY 60 ----- 60
 Db 1318 AACGAGTCGGAATGCAAGTGCAGTCGGGATGCGGATGCGGATGAGACAAAACATGATCATTTGGA 1377
 QY 60 ----- 60
 Db 1378 GCGACCCAGTTAGCGCCACATGTCGGGATTCGGTTCCTTCCCAATAAGCCGCTCAATGGC 1437
 QY 61 -----AsnSe 62
 Db 1438 CCATGAATATGCATATTTGATACGAGATGACAGCGTATCATGGGCGAGCAATGAATC 1497
 QY 62 rIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGlyAlaG1 82
 Db 1498 CATACGTCACAATCTGTCGCTGCACAAACCGCTTTATGAGGGTCCAAAACGAGGCGACCGG 1557
 QY 82 yLysSerSerTrpTrpValIleAsnProAspAlaLysProGlyMetAsnProArgArg 101
 Db 1558 CAAGTCATCTCTGGTGGATGCTCAACCCGGAGGCCAAGCCCGGCAAGTCTGTGCGCCGC 1615
 RESULT 14
 ABL05684
 ID ABL05684 standard; cDNA; 25891 BP.
 AC ABL05684;
 XX 26-MAR-2002 (first entry)
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 11534.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical; gene; ss.
 XX Drosophila melanogaster.
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 XX 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX P-PSDB; ABB61581.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX Claim 1; SEQ ID NO 11534; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABBS72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 25891 BP; 7426 A; 5074 C; 5058 G; 8333 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.21e-22 Length: 25891
 Score: 262.00 Matches: 47
 Percent Similarity: 91.94% Conservative: 10
 Best Local Similarity: 75.81% Mismatches: 5
 Query Match: 47.21% Indels: 0
 DB: 4 Gaps: 0

US-09-844-353A-54 (1-103) x ABL05684 (1-25891)

QY 1 LysLysThrThrThrArgAGAsnAlaTIPGLyAsnMetSerTyrAlaGluLeuIleThr 20
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 11002 AAGAGAACTCAACGGCTGCATGCGATGGGAAATCTCTTAAGCGATCTCATCAGC 11061
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTIPMet 40
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 11062 CATGCCATTGGATCGGCCACCGACAAACGATTGACACTGAGTCAGATTACGATGGATG 11121
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerIleAsnSerAlaGlyTIPLys 60
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 11122 GTCAGAAATGGCATATTTCAAAGGACAAAGCGGATTCGAATAGCAGTCGCGGATGGAAG 11181
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 61 AsnSer 62
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 11182 GTGAGT 11187
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 15
 AAD55840
 ID AAD55840 standard; cDNA; 5123 BP.
 XX AC AAD55840;
 XX DT 07-AUG-2003 (first entry)
 XX DE Human nucleic acid associated protein (NAAP)-11 cDNA.
 XX KW Human; nucleic acid associated protein; NAAP; stroke; AIDS; nootropic;
 KW cancer; atherosclerosis; neurological; epilepsy; Huntington's disease;
 KW developmental disorder; antiinflammatory; neuroprotective; thyromimetic;
 KW Cushing's syndrome; infection; gene therapy; cytostatic; anticonvulsant;
 KW cerebroprotective; hypothyroidism; cell proliferative disorder; allergy;
 KW gene; ss.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT CDS 179..1945
 FT FT /*tag= a
 FT FT /product= "Human NAAP protein"
 XX PN WO2003006618-A2.
 XX PD 23-JAN-2003.
 XX PF 10-JUL-2002; 2002WO-US021971.
 XX PR 12-JUL-2001; 2001US-0305089P.
 PR 12-JUL-2001; 2001US-0305104P.
 PR 13-JUL-2001; 2001US-0305325P.
 PR 13-JUL-2001; 2001US-0305390P.
 PR 19-JUL-2001; 2001US-0306960P.
 PR 20-JUL-2001; 2001US-0306694P.
 PR 27-JUL-2001; 2001US-0308170P.
 XX PA (INCY-) INCYTE GENOMICS INC.

XX Swarnakar A, Richardson TW, Warren BA, Griffin JA, Tang YT;
 PI Yue H, Baughn MR, Emerling BM, Lal PG, Lu DAM, Forsythe IU;
 PI Ramkumar J, Li JX, Becha SD, Duggan BM, Sanjanwala MM, Lee EA;
 PI Burford N, Elliott VS, Ison CH, Ding L, Borowsky ML, Yao MG;
 PI Barroso I, Tran B, Wallia NK, Hafalia AJA, Nguyen DB, Lu Y;
 PI Arvizu CS;
 XX WPI; 2003-221732/21.
 DR P-PSDB; AAE37024.
 XX New human nucleic acid associated proteins (NAAP), useful for diagnosing,
 PT treating and preventing diseases or conditions associated with the
 PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
 PT infections.
 XX Claim 5; Page 239-241; 260pp; English.

XX The invention relates to human nucleic acid associated proteins (NAAP)
 CC and their corresponding nucleic acid sequences. The invention is useful
 CC in diagnosing, treating and preventing diseases or conditions associated
 CC with the decreased expression or overexpression of NAAP, such as cell
 CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
 CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
 CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
 CC disorders, or infections. It is also useful in assessing the effects of
 CC exogenous compounds on the expression of nucleic acid and amino acid
 CC sequences of NAAP. The NAAP or its fragments are useful in screening
 CC compounds for which acts as their agonist or antagonist. The microarray
 CC is useful in monitoring or measuring protein-protein interactions, drug-
 CC target interactions, and gene expression profiles. NAAP DNA is used in
 CC gene therapy. The present sequence is human NAAP cDNA

SQ Sequence 5123 BP; 1383 A; 1103 C; 1115 G; 1522 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2.09e-18 Length: 5123
 Score: 222.50 Matches: 45
 Percent Similarity: 66.67% Conservative: 19
 Best Local Similarity: 46.88% Mismatches: 21
 Query Match: 40.09% Indels: 11
 DB: 7 Gaps: 4

US-09-844-353A-54 (1-103) x AAD55840 (1-5123)

QY 14 SerTyrAlaGluLeuIleThrThrAlaIleMetAlaSerProGluLysArgLeuThrLeu 33
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 422 AGTTATGCCAGCTCTTACATTGTCATTTGCAATTATAGTCCACCCAAAAGAAAATGACTTTA 481
 ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 34 AlaGlnValTyrGluTIPMetValGlnAsnValProTyrPheArgAspLysGlyAspSer 53
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 482 AGTGAATTTATCAGTGGATTTGTGATACTTCCCATATTATAGAGAGGTGGC----- 535
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 54 AsnSerSerAlaGlyTIPLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPhe 73
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 536 -----AGTGGTTGGAAGAATTCATACGACATAATCTGTCTATGACCAATGTTTC 586
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 74 MetArgIle-----GlnAsnGluGlyAlaGlyLysSerSerTIPTrpValIle----- 89
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 587 CTTAAAGTCGCTCGATCTTAAGGATGACCTCGAAGGGGTCTCTACTGGGCATAGACACC 646
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 90 AsnPro-----AspAlaLysProGlyMetAsnProArgArgThrArg 103
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 647 AATCCGAGGAAGATGCGCTGCTACTCGCCAAAGAGAGAGGCGCAGCA 694
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Search completed: July 29, 2004, 10:40:16
 Job time : 350 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: July 29, 2004, 10:27:35 ; Search time 81 Seconds
(without alignments)
705.678 Million cell updates/sec

Title: US-09-844-353A-54
Perfect score: 555
Sequence: 1 KKTTRNAGNMSYAEIIT.....SSWWVINDAKPGMPPRTR 103

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USFTO_spool_p/US0984433/runat_23072004_084914_28266/app_query.fasta_1.263
-DB=issued_Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCF=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09844333 @CGN 1.128 @runat_23072004_084914_28266 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	555	100.0	3499	US-08-857-076-43	Sequence 43, Appl
2	410.5	74.0	1073	US-09-645-629-22	Sequence 22, Appl
3	410.5	74.0	3394	US-09-620-312D-314	Sequence 314, App
4	378.5	68.2	2704	US-08-857-076-44	Sequence 44, Appl
5	185.5	33.4	2830	US-07-882-292-1	Sequence 1, Appli
6	185.5	33.4	2830	US-08-331-644-1	Sequence 1, Appli
7	185.5	33.4	2830	PCT-US93-04102-1	Sequence 1, Appli
8	184.5	33.2	1860	US-08-331-644-3	Sequence 3, Appli
9	184.5	33.2	1860	PCT-US93-04102-3	Sequence 3, Appli
10	182.5	32.9	1659	US-09-083-352-3	Sequence 3, Appli
11	182.5	32.9	1659	US-09-083-351-3	Sequence 3, Appli
12	182.5	32.9	3946	US-09-083-351-1	Sequence 1, Appli

13	182.5	32.9	3946	3	US-09-083-352-1	Sequence 1, Appli
14	182	32.8	3342	2	US-08-742-753-3	Sequence 3, Appli
15	181.5	32.7	1155	2	US-08-331-644-4	Sequence 4, Appli
16	181.5	32.7	1155	5	PCT-US93-04102-4	Sequence 4, Appli
17	178.5	32.2	2517	1	US-07-906-930E-1	Sequence 1, Appli
18	178.5	32.2	3465	4	US-09-023-655-1136	Sequence 1136, Ap
19	176.5	31.8	2939	1	US-07-906-930E-3	Sequence 3, Appli
20	176	31.7	2160	4	US-09-372-668-1	Sequence 1, Appli
21	176	31.7	2160	4	US-09-697-377-1	Sequence 1, Appli
22	176	31.7	2160	4	US-09-696-868-1	Sequence 1, Appli
23	176	31.7	2160	4	US-09-697-341-1	Sequence 1, Appli
24	175	31.5	1869	4	US-09-372-668-3	Sequence 3, Appli
25	175	31.5	1869	4	US-09-697-377-3	Sequence 3, Appli
26	175	31.5	1869	4	US-09-696-868-3	Sequence 3, Appli
27	175	31.5	1869	4	US-09-697-341-3	Sequence 3, Appli
28	173.5	31.3	1965	4	US-09-220-132-21	Sequence 21, Appl
29	170.5	30.7	232	4	US-09-016-434-595	Sequence 595, App
30	166	29.9	1766	4	US-09-833-381-1265	Sequence 1265, Ap
31	164.5	29.6	785	1	US-08-329-055-5	Sequence 5, Appli
32	164.5	29.6	3441	2	US-08-742-753-1	Sequence 1, Appli
33	163.5	29.5	620	1	US-08-329-055-2	Sequence 2, Appli
34	163.5	29.5	5080	4	US-09-976-594-495	Sequence 495, App
35	160.5	28.9	1309	4	US-09-976-594-927	Sequence 927, App
36	150	27.0	493	4	US-09-833-381-76	Sequence 76, Appl
37	137.5	24.8	715	3	US-08-998-416-648	Sequence 648, App
38	134	24.1	229	4	US-09-016-434-814	Sequence 814, App
39	133	24.0	167	4	US-09-833-381-467	Sequence 467, App
40	130	23.4	1634	4	US-09-087-134-13	Sequence 13, Appl
41	130	23.4	1793	3	US-09-113-309-1	Sequence 1, Appli
42	130	23.4	1793	3	US-09-521-109-1	Sequence 1, Appli
43	130	23.4	1793	4	US-09-562-332-1	Sequence 1, Appli
44	124.5	22.4	1605	4	US-09-087-134-10	Sequence 10, Appl
45	124	22.3	1668	4	US-09-087-134-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-857-076-43
; Sequence 43, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 3499
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-08-857-076-43

Alignment Scores:
Pred. No.: 2,42e-62 Length: 3499
Score: 555.00 Matches: 103
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-844-353A-54 (1-103) x US-08-857-076-43 (1-3499)

QY	1	LysLysThrThrThrArgArgAsnAlaTrrGlyAsnMetSerTyrAlaGluLeuIleThr	20
Db	727	AAAAAGACAACACGACGACGACCGCTTGGGAAATATGTATCATATGTAACATCATCT	786
QY	21	ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrrMet	40
Db	787	ACAGCCATTATGGCTAGTCAGAGAAACGGTTAACTCTTGACAAAGTTTACGAATGGATG	846
QY	41	ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrrLys	60
Db	847	GTCAGAAATGTTCCATATCTCAGGATAAGGAGATTCGAACAGTTCAGCTGGATGAAG	906
QY	61	AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly	80
Db	907	AACTCGATCCGTCACATCTCTCTTCATTCCTGTTTCATCGGAATTCAGAAATGAAGGA	966
QY	81	AlaGlyLysSerSerTrrTrrValIleAsnProAspAlaLysProGlyMetAsnProArg	100
Db	967	GCCGGAAGAGCTCTGTGTGGTTATTAATCCAGATGCAGGAGGAGGAGGAGGAGGAGG	1026
QY	101	ArgThrArg	103
Db	1027	CGTACACGT	1035
RESULT 2			
US-09-645-629-22			
; Sequence 22, Application US/09645629			
; Patent No. 6472515			
; GENERAL INFORMATION:			
; APPLICANT: Climent-Johansson, Isabel			
; APPLICANT: Dahlman-Wright, Karin			
; APPLICANT: Lake, Staffan			
; APPLICANT: Wasserman, Wyeth			
; TITLE OF INVENTION: NOVEL RESPONSE ELEMENT			
; FILE REFERENCE: 13425-032001			
; CURRENT APPLICATION NUMBER: US/09/645, 629			
; CURRENT FILING DATE: 2000-08-24			
; PRIOR APPLICATION NUMBER: SE 9904269-9			
; PRIOR FILING DATE: 1999-11-25			
; PRIOR APPLICATION NUMBER: US 60/151,967			
; PRIOR FILING DATE: 1999-08-31			
; PRIOR APPLICATION NUMBER: SE 9903009-0			
; PRIOR FILING DATE: 1999-08-26			
; NUMBER OF SEQ ID NOS: 34			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 22			
; LENGTH: 1073			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: synthetically generated GST-AfxDBD construct			
US-09-645-629-22			
Alignment Scores:			
Pred. No.:	3,4e-44	Length:	1073
Score:	410.50	Matches:	73
Percent Similarity:	86.27%	Conservative:	15
Best Local Similarity:	71.57%	Mismatches:	13
Query Match:	73.96%	Indels:	1
DB:	4	Gaps:	1
US-09-844-353A-54 (1-103) x US-09-645-629-22 (1-1073)			
QY	1	LysLysThrThrThrArgArgAsnAlaTrrGlyAsnMetSerTyrAlaGluLeuIleThr	20
Db	715	CGAAGGGAGGCTCCCGCGGATGCTGGGAAATCAGTCATATCGAGAACTCATCAGC	774
QY	21	ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrrMet	40
Db	775	CAGGCCATTGAAGCGCCCGGAGAGCGACTGACATCTGCCAGATCTACGATGGATG	834
QY	41	ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrrLys	60
US-09-620-312D-314			
; Sequence 314, Application US/09620312D			
; Patent No. 6569662			
; GENERAL INFORMATION:			
; APPLICANT: Tang, Y. Tom			
; APPLICANT: Liu, Chenghua			
; APPLICANT: Asundi, Vinod			
; APPLICANT: Zhang, Jie			
; APPLICANT: Ren, Feiyan			
; APPLICANT: Chen, Rui-hong			
; APPLICANT: Zhao, Qing A.			
; APPLICANT: Wehrman, Tom			
; APPLICANT: Xue, Aidong J.			
; APPLICANT: Yang, Yonghong			
; APPLICANT: Wang, Jian-Rui			
; APPLICANT: Zhou, Ping			
; APPLICANT: Ma, Yunqing			
; APPLICANT: Wang, Dunrui			
; APPLICANT: Wang, Zhiwei			
; APPLICANT: John Tillinghast			
; APPLICANT: Drmanac, Radolje T.			
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and			
; TITLE OF INVENTION: Polypeptides			
; FILE REFERENCE: 784CIP2B			
; CURRENT APPLICATION NUMBER: US/09/620,312D			
; CURRENT FILING DATE: 2000-07-19			
; PRIOR APPLICATION NUMBER: 09/552,317			
; PRIOR FILING DATE: 2000-04-25			
; PRIOR APPLICATION NUMBER: 09/488,725			
; PRIOR FILING DATE: 2000-01-21			
; NUMBER OF SEQ ID NOS: 1105			
; SOFTWARE: pt_FL_genes Version 1.0			
; SEQ ID NO 314			
; LENGTH: 3394			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (417)..(1934)			
US-09-620-312D-314			
Alignment Scores:			
Pred. No.:	1,71e-43	Length:	3394
Score:	410.50	Matches:	73
Percent Similarity:	86.27%	Conservative:	15
Best Local Similarity:	71.57%	Mismatches:	13
Query Match:	73.96%	Indels:	1
DB:	4	Gaps:	1
US-09-844-353A-54 (1-103) x US-09-620-312D-314 (1-3394)			
QY	1	LysLysThrThrThrArgArgAsnAlaTrrGlyAsnMetSerTyrAlaGluLeuIleThr	20
Db	690	CGAAGGGAGGCTCCCGCGGATGCTGGGAAATCAGTCATATCGAGAACTCATCAGC	749
QY	21	ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrrMet	40
Db	750	CAGGCCATTGAAGCGCCCGGAGAGCGACTGACATCTGCCAGATCTACGATGGATG	809

QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerAlaGlyTrpLys 60
Db 810 GTCCGTACTGTACCTCTACTTCAAGACAAGGGTGACACAGCTCAGCAGGATGGAAG 869
QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db 870 AACTCGATCCGCCACACCTGTCCCTGCACAGCAAGTTTCATCAAGGTTTCACACAGGACC 929
QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db 930 ACCGGCAAAAGCTCTGTGTGATGCTGACCTCAGGAGGAGCGGCAAGAGCCGCC 989
QY 100 ArgArg 101
Db 990 CGCCGC 995
RESULT 4
US-08-857-076-44
; Sequence 44, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 2704
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-08-857-076-44
Alignment Scores:
Pred. No.: 1,87e-39 Length: 2704
Score: 378.50 Matches: 67
Percent Similarity: 75.93% Conservative: 15
Best Local Similarity: 62.04% Mismatches: 21
Query Match: 68.20% Indels: 5
DB: 3 Gaps: 1
US-09-844-353A-54 (1-103) x US-08-857-076-44 (1-2704)
QY 1 LysLysThrThr-----ThrArgAsnAlaTrpGlyAsnMetSerTyr 15
Db 398 AAGAAGCCACCGATCAATTTGGCACAGAAACCGATCCATGGGTGAGGATCTAT 457
QY 16 AlaGluLeuIleThrThrAlaIleMetAlaSerProGlyLysArgLeuThrLeuAlaGln 35
Db 458 TCGGATATCATTTGCCAAGCAATTTGAATCGGCCACAGCGAAGGCTTAACTCAATGAG 517
QY 36 ValTyrGluTrpMetValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSer 55
Db 518 ATTATCAATGGTCTCTCTGATATATATCTCTTTGGAGAACGATCTAGTCCCGAGGAG 577
QY 56 SerAlaGlyTrpLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArg 75
Db 578 GCGCCGGATGGAGAACTCGATCCCTCACAATCTCTCTTCATCTCTGTTTCATGCCA 637
QY 76 IleGlnAsnGluGlyAlaGlyLysSerSerTrpTrpValIleAsnProAspAlaLysPro 95
Db 638 ATTCAAGATGAAGAGCGGAGAGAGCTGCTGTGTGGTTTATTATCCAGATGCAAGGCCA 697

QY 96 GlyMetAsnProArgArgThrArg 103
Db 698 GGAATGAATCCACGGCGTACACGT 721
RESULT 5
US-07-882-292-1
; Sequence 1, Application US/07882292
; Patent No. 5324638
; GENERAL INFORMATION:
; APPLICANT: Tao, Wufan
; APPLICANT: Lai, Eseng
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: c/o Cooper and Dunham, 30 Rockefeller
; STREET: Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/882,292
; FILING DATE: 19920513
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41472
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-977-9550
; TELEFAX: 212-664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2830 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 443..1882
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 926..1255
; OTHER INFORMATION: /note= "nucleotide sequence encoding DNA
; OTHER INFORMATION: binding domain homology"
; FEATURE:
; NAME/KEY: misc signal
; LOCATION: 1883..1885
; OTHER INFORMATION: /note= "translation termination codon"
US-07-882-292-1
Alignment Scores:
Pred. No.: 3.16e-14 Length: 2830
Score: 185.50 Matches: 35
Percent Similarity: 68.29% Conservative: 21
Best Local Similarity: 42.68% Mismatches: 19
Query Match: 33.42% Indels: 7
DB: 1 Gaps: 2
US-09-844-353A-54 (1-103) x US-07-882-292-1 (1-2830)

TELEFAX: 212-664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2830 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 443..1882
OTHER INFORMATION:
FEATURE:
NAME/KEY: misc_feature
LOCATION: 926..1255
OTHER INFORMATION: /note= "nucleotide sequence encoding DNA
OTHER INFORMATION: binding domain homology"
FEATURE:
NAME/KEY: misc_signal
LOCATION: 1883..1885
OTHER INFORMATION: /note= "translation termination codon"
PCT-US93-04102-1

Alignment Scores:
Pred. No.: 3.16e-14 Length: 2830
Score: 185.50 Matches: 35
Percent Similarity: 68.29% Conservative: 21
Best Local Similarity: 42.68% Mismatches: 19
Query Match: 33.42% Indels: 7
DB: 5 Gaps: 2

US-09-844-353A-54 (1-103) x PCT-US93-04102-1 (1-2830)

QY 14 SerTyAlaGluLeuIleThrAlaIleMetAlaSerProGluLysArgLeuThrLeu 33
Db 968 ACCTACACCGCGTCATCATGATGGCCATCAGGAGAGTCCCGAGAGAGCCCTGACGCTC 1027
QY 34 AlaGlnValTyrGluTrpMetValGlnAsnValProTyrPheArgAspLysGlyAspSer 53
Db 1028 AACGGCATCTACGAGTTCATCATGAGAACTCCCTTACTACCGCGAG----- 1075
QY 54 AsnSerSerAlaGlyTrpLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPhe 73
Db 1076 ---AACAGCAGGCGTGGCAGAACTCCATCCGCCACAACCTGCCCTCAACAAGTGCTTC 1132
QY 74 MetArgIleGlnAsn-----GluGlyAlaGlyLysSerSerTrpTrpValIleAsnPro 91
Db 1133 GTGAAGGTACCGCGCCACTAGCAGACCGCGGCAAGGGCAACTACTGATGCTGGACCCG 1192
QY 92 AspAla 93
Db 1193 TCAGAGC 1198

RESULT 8

US-08-331-644-3
Sequence 3, Application US/08331644
Patent No. 5976872
GENERAL INFORMATION:
APPLICANT: Tao, Wufan
APPLICANT: Lai, Eseng
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/331,644
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882,292
FILING DATE: 13-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41472-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1860 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-331-644-3
Alignment Scores:
Pred. No.: 2.37e-14 Length: 1860
Score: 184.50 Matches: 36
Percent Similarity: 67.07% Conservative: 19
Best Local Similarity: 43.90% Mismatches: 20
Query Match: 33.24% Indels: 7
DB: 2 Gaps: 2
US-09-844-353A-54 (1-103) x US-08-331-644-3 (1-1860)
QY 14 SerTyAlaGluLeuIleThrAlaIleMetAlaSerProGluLysArgLeuThrLeu 33
Db 765 TCGTATACGCGCTCATCATCCATCGCCATCTCTCAGAGCCCAAGAGCCCTGACGCTC 824
QY 34 AlaGlnValTyrGluTrpMetValGlnAsnValProTyrPheArgAspLysGlyAspSer 53
Db 825 AGCGATTCGCGAGTTCATCATCAGCAGCGCTTCCCTTACTACCGGAGAG----- 875
QY 54 AsnSerSerAlaGlyTrpLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPhe 73
Db 876 -----TTCCCGCTGGCAGAGACAGCATCCGTCACAACCTGCTGCTCAACGACTGCTTC 929
QY 74 MetArgIleGlnAsnGlu-----GlyAlaGlyLysSerSerTrpTrpValIleAsnPro 91
Db 930 GTCAAGATCCCGCGCAACCGGCAACCGGCAAGGGCAACTACTGAGCTCGACCCG 989
QY 92 AspAla 93
Db 990 GAGTCC 995
RESULT 9
PCT-US93-04102-3
Sequence 3, Application PC/TUS9304102
GENERAL INFORMATION:
APPLICANT: Tao, Wufan
APPLICANT: Lai, Eseng
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA

; ZIP: 10112
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/04102
 ; FILING DATE: 19930430
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/882,292
 ; FILING DATE: 13-MAY-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 41472A-PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-977-9550
 ; TELEFAX: 212-664-0525
 ; TELEX: 422523 COOP UI
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1860 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; PCT-US93-04102-3

Alignment Scores:
 Pred. No.: 2,37e-14 Length: 1860
 Score: 184.50 Matches: 36
 Percent Similarity: 67.0% Conservative: 19
 Best Local Similarity: 43.90% Mismatches: 20
 Query Match: 33.24% Indels: 7
 DB: 5 Gaps: 2

US-09-844-353A-54 (1-103) x PCT-US93-04102-3 (1-1860)

QY 14 SerTyrAlaGluLeuThrAlaMetAlaSerProGluLysArgLeuThrLeu 33
 Db 765 TCGTACATCGGCTCATCACCATGGCCATCTCGCAGAGCCGACGAGCGCTGACGTC 824
 QY 34 AlaGlnValTyrGluTrpMetValGlnAsnValProTyrPheArgAspLysGlyAspSer 53
 Db 825 ACGGAGATCTGGAGTTCATCAGACGCGCTTCCCTTACTACCGGGAAG----- 875
 QY 54 AsnSerSerAlaGlyTrpLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPhe 73
 Db 876 -----TTCCCGCTTGGCAGAACAGCATCGTCACACCTCGCTCAACGACTGCTTC 929
 QY 74 MetArgIleGlnAsnGlu-----GlyAlaGlyLysSerSerTrpTrpValIleAsnPro 91
 Db 930 GTCAGATCCCGCGCAACCGGCAACCGGCGCAAGGCACTACTGACGCTGACCCG 989
 QY 92 AspAla 93
 Db 990 GAGTCC 995

RESULT 10
 US-09-083-351-3
 ; Sequence 3, Application US/09083351
 ; Patent No. 6087107
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheffield, Val C.
 ; APPLICANT: Alward, Wallace L.M.
 ; APPLICANT: Stone, Edwin M.
 ; APPLICANT: Nishimura, Darryl
 ; APPLICANT: Patil, Shiva
 ; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
 ; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
 ; TITLE OF INVENTION: TRANSCRIPTION FACTOR

; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
 ; STREET: One Post Office Square
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109-2170
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/083,351
 ; FILING DATE: 22-MAY-1998
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Arnold, Beth E.
 ; REGISTRATION NUMBER: 35,430
 ; REFERENCE/DOCKET NUMBER: UIA-029.02
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-832-1000
 ; TELEFAX: 617-832-7000
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1659 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-09-083-351-3

Alignment Scores:
 Pred. No.: 3,68e-14 Length: 1659
 Score: 182.50 Matches: 33
 Percent Similarity: 68.29% Conservative: 23
 Best Local Similarity: 40.24% Mismatches: 19
 Query Match: 32.88% Indels: 7
 DB: 3 Gaps: 2

US-09-844-353A-54 (1-103) x US-09-083-351-3 (1-1659)
 QY 14 SerTyrAlaGluLeuThrAlaMetAlaSerProGluLysArgLeuThrLeu 33
 Db 244 AGTACATCGGCTCATCACCATGGCCATCCAGAACGCCGACAGAGATCACCTG 303
 QY 34 AlaGlnValTyrGluTrpMetValGlnAsnValProTyrPheArgAspLysGlyAspSer 53
 Db 304 AACGCATCTACCGATTCATCATGACGCGCTTCCCTTCTACCGGGAC----- 351
 QY 54 AsnSerSerAlaGlyTrpLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPhe 73
 Db 352 ---AACAGCAGGCTGGCAGAACAGCATCGCCACACCTCTCGCTCAACGAGTGTTC 408
 QY 74 MetArgIle-----GlnAsnGluGlyAlaGlyLysSerSerTrpTrpValIleAsnPro 91
 Db 409 GTCAGGTGCGCGCGACGACAGAACGCCGCGGAGGCGAGCTACTGGACGCTGACCG 468
 QY 92 AspAla 93
 Db 469 GACTCC 474

RESULT 11
 US-09-083-352-3
 ; Sequence 3, Application US/09083352
 ; Patent No. 6207450
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheffield, Val C.
 ; APPLICANT: Alward, Wallace L.M.
 ; APPLICANT: Stone, Edwin M.
 ; APPLICANT: Nishimura, Darryl
 ; APPLICANT: Patil, Shiva


```
; Sequence 1, Application US/09083352
; Patent No. 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,352
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UTA-029.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 475..2133
US-09-083-352-1

Alignment Scores:
Pred. No.: 1.24e-13 Length: 3946
Score: 182.50 Matches: 33
Percent Similarity: 68.29% Conservative: 23
Best Local Similarity: 40.24% Mismatches: 19
Query Match: 32.88% Indels: 7
DB: 3 Gaps: 2

US-09-844-353A-54 (1-103) x US-09-083-352-1 (1-3946)
QY 14 SerTyrAlaGluLeuThrAlaIleMetAlaSerProGluLysArgLeuThrLeu 33
DB 718 AGCTACATCGCGCTCATCCATGCGCATCCAGACGCCCGGCAAGAGATCACCCGTG 777
QY 34 AlaGlnValTyrGluTrpMetValGlnAsnValProTyrPheArgAspLysGlyAspSer 53
DB 778 AACGGCATCTACCGTTCATCATGACCGCTTCCCTCTACCGGAC-----825
QY 54 AsnSerSerAlaGlyTrpLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPhe 73
DB 826 ---AACACAGCAGGGCTGGCAGAACAGCATCCGCCCAACCTCTCGCTCAACGAGTGCTTC 882
QY 74 MetArgIle-----GlnAsnGluGlyAlaGlyLysSerSerTyrTrpValIleAsnPro 91
DB 883 GTCAAGGTCCCGCGGACGACAGAACGCGGCAAGGCGGACGACTACTGGACGCTGCACCG 942
QY 92 AspAla 93

; Sequence 1, Application US/08742753
; Patent No. 5861278
; GENERAL INFORMATION:
; APPLICANT: WONG, Gordon G.
; APPLICANT: YAO, Kwok-Ming
; TITLE OF INVENTION: HNF3-delta Compositions
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,753
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: LAZAR, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5277
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8260
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3342 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 70..2358
US-08-742-753-3

Alignment Scores:
Pred. No.: 1.14e-13 Length: 3342
Score: 182.00 Matches: 38
Percent Similarity: 56.52% Conservative: 14
Best Local Similarity: 41.30% Mismatches: 28
Query Match: 32.79% Indels: 12
DB: 2 Gaps: 3

US-09-844-353A-54 (1-103) x US-08-742-753-3 (1-3342)
QY 9 AlaTrpGlyAsn-----MetSerTyrAlaGluLeuIleThrThr 21
DB 751 TCCTGGCAGAACTGTGTCTGTGAGGGCCACCTTACTTCTATGCGCATGATACAATTC 810
QY 22 AlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMetVal 41
DB 811 GCCATCAACAGCAGCTGAGGAGGCGCATGCTTTTGAAGACATCTATACGTGGATTGAG 870
QY 42 GlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLysAsn 51
DB 871 GACCATTTCCCTACTCTTAAG-----CACATTGCCAACGCCAGGCTGGAGAAC 918
QY 62 SerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGlyAla 81
DB 919 TCCATCCGCCAACACTTTTCCCTGCGACATGTTGTTCGGG---GAGACGCTGCCAAT 975
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 29, 2004, 11:31:55 ; Search time 358 Seconds
(without alignments)
1410.680 Million cell updates/sec

Title: US-09-844-353A-54

Perfect score: 555

Sequence: 1 KKTTRRNAGNNSYAELIT.....SSWVNPDAKPGMNPRT 103

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3222919 seqs, 2451570024 residues

Total number of hits satisfying chosen parameters: 6445838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09844353/runat 23072004 084915 28361/app query.fasta_1.263
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonsum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09844353 @CGN 1.1 912 @runat 23072004 084915 28361
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

ALIGNMENTS

RESULT 1

US-09-205-658-43

; Sequence 43, Application US/09205658

; Patent No. US20010029617A1

; GENERAL INFORMATION:

; APPLICANT: Ruvkun, Gary

; APPLICANT: OSG, Scott

; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR

; IMPAIRED GLUCOSE TOLERANCE CONDITIONS

; FILE REFERENCE: 00786/351004

; CURRENT APPLICATION NUMBER: US/09/205,658

; CURRENT FILING DATE: 1998-12-03

; EARLIER APPLICATION NUMBER: 08/857,076

; EARLIER FILING DATE: 1997-05-15

; EARLIER APPLICATION NUMBER: 08/888,534

; EARLIER FILING DATE: 1997-07-07

; EARLIER APPLICATION NUMBER: US98/10080

; EARLIER FILING DATE: 1998-05-15

; NUMBER OF SEQ ID NOS: 328

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 43

; LENGTH: 3499

; TYPE: DNA

Sequence 43, Appl
Sequence 43, Appl
Sequence 43, Appl
Sequence 245, App
Sequence 134, App
Sequence 52, Appl
Sequence 4, Appl
Sequence 11, Appl
Sequence 22, Appl
Sequence 314, App
Sequence 314, App
Sequence 44, Appl
Sequence 44, Appl
Sequence 44, Appl
Sequence 325, App
Sequence 743, App
Sequence 1132, App
Sequence 1132, App
Sequence 24, Appl
Sequence 589, App
Sequence 20688, A
Sequence 1, Appl
Sequence 611, App
Sequence 611, App
Sequence 133, App
Sequence 1045, App
Sequence 1045, App
Sequence 135, App
Sequence 802, App
Sequence 1, Appl
Sequence 1, Appl
Sequence 304, App
Sequence 336, App
Sequence 78, Appl
Sequence 712, App
Sequence 1820, App
Sequence 1887, App
Sequence 1887, App
Sequence 110, App
Sequence 337, App
Sequence 9, Appl
Sequence 4, Appl
Sequence 11, Appl
Sequence 6, Appl

Db 847 GTCCAGATGTCATACCTTCAGGATACGGAGATTCGACAGTTCAGCTGGATGAAG 906
QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db 907 AACTCGATCGCTCACAATCTGTCTCTTCATTCCTGTTTCATGCGAATTCAGATGAAGGA 966
QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAspAlaLysProGlyMetAsnProArg 100
Db 967 GCCGGAAGAGCTCGTGGTGGGTATTATATCCAGATCAAAAGCCAGGAATGAATCCACGG 1026
QY 101 ArgThrArg 103
Db 1027 CGTACACGT 1035

RESULT 4

US-10-717-597-245
; Sequence 245, Application US/10717597
; Publication No. US20040110221A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Burczynski, Michael E.
; APPLICANT: Twine, Natalie C.
; APPLICANT: Dörner, Andrew J.
; APPLICANT: Trepicchio, William L.
; APPLICANT: Slonim, Donna K.
; APPLICANT: Stover, Jennifer A.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND OTHER SOLID TUMORS
; FILE REFERENCE: AM101080L
; CURRENT APPLICATION NUMBER: US/10/717,597
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US 60/459,782
; PRIOR FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: US 60/427,982
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 4904
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 245
; LENGTH: 3183
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-717-597-245

Alignment Scores:
Pred. No.: 3,35e-53 Length: 3183
Score: 440.50 Matches: 79
Percent Similarity: 90.20% Conservative: 13
Best Local Similarity: 77.45% Mismatches: 9
Query Match: 79.37% Indels: 1
DB: 17 Gaps: 1

US-09-844-353A-54 (1-103) x US-10-717-597-245 (1-3183)

QY 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
Db 1366 AGGAATGTTCTGTCGCGGGGAAAGCGCTCGGGAACCTGTCTTACCGGACCTGATCACC 1425
QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
Db 1426 CGCGCCATCGAGAGCTCCCGGCAAAACGCTCACTCTGTCCAGATCTACGAGTGGATG 1485
QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db 1486 GTGCGTGGTGCCTTCTTCAAGGATAGGGGAGAGGAGCAGCAACAGCTCTCGCGGCTCGAAG 1545
QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db 1546 AACTCCATCGGCAACACCTGTCTCATGTGATGCTATGCTGCGGTCCAGATGAGGGA 1605
QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db 1606 ACTGGCAAGAGCTCTTGGTGGATCATCAACCTGTATGGGGGAGAGCGGAAAGCCCC 1665

QY 100 ArgArg 101
Db 1666 CGGCGG 1671

RESULT 5

US-10-007-926A-134
; Sequence 134, Application US/10007926A
; Publication No. US20030143539A1
; GENERAL INFORMATION:
; APPLICANT: BERTUCCI, FRANCOIS
; APPLICANT: HOULGATTE, REMI
; APPLICANT: BIRNEAUM, DANIEL
; APPLICANT: NGUYEN, CATHERINE
; APPLICANT: VIENS, PATRICE
; APPLICANT: FERT, VINCENT
; TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
; FILE REFERENCE: 1546-R-00
; CURRENT APPLICATION NUMBER: US/10/007,926A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/254,090
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 134
; LENGTH: 5723
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: forkhead box ola (rhabdomyosarcoma) (FOXO1A)
; OTHER INFORMATION: gene.
US-10-007-926A-134

Alignment Scores:
Pred. No.: 5,96e-51 Length: 5723
Score: 427.50 Matches: 75
Percent Similarity: 91.18% Conservative: 18
Best Local Similarity: 73.53% Mismatches: 8
Query Match: 77.03% Indels: 1
DB: 15 Gaps: 1

US-09-844-353A-54 (1-103) x US-10-007-926A-134 (1-5723)

QY 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
Db 836 AAGAGCAGCTCGTCCCGGCAACCGTGGGGCAACCTGTCTTACCGCCAGCTCATCACC 895
QY 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
Db 896 AAGCCATCGAGAGCTCGGCGGAGAGCGGCTCACGCTGCGCAGATCTACGAGTGGATG 955
QY 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db 956 GTCAAGAGCGTGCCTTACTTCAAGGATAGGGTGACAGCAACAGCTCGGCGGCTGGAAG 1015
QY 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db 1016 AATTCAATTCGTATAATCTGTCCCTACACAGCAAGTTCATTCGTGCGAATGAAGGA 1075
QY 81 AlaGlyLysSerSerTrpTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db 1076 ACTGGAAAAGTCTTGGTGGATGCTCAATCCAGAGGGTGGCAAGAGCGGAAATCTCCT 1135
QY 100 ArgArg 101
Db 1136 AGGAGA 1141

RESULT 6

US-10-341-434-52
; Sequence 52, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies

```
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 90 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52
; LENGTH: 5723
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (386)..(2350)
; OTHER INFORMATION:
US-10-341-434-52

Alignment Scores:
Pred. No.: 5,96e-51 Length: 5723
Score: 427.50 Matches: 75
Percent Similarity: 91.18% Conservative: 18
Best Local Similarity: 73.53% Mismatches: 8
Query Match: 77.03% Indels: 1
DB: 16 Gaps: 1

US-09-844-353A-54 (1-103) x US-10-341-434-52 (1-5723)
Qy 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
Db 836 AAGAGCAGCTCGTCCCGCCGCAACGCGTGGGGCAACCTGTCTACGCGGACCTCATCACC 895
Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
Db 896 AAGGCCATCGAGAGCTCGCGGAGAGCGGCTCACGCTGTGCGCAGATCTACGAGTGGATG 955
Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db 956 GTCGAAGAGCGTCCCTACTTCAAGGATAAGGGTGACAGCAACAGCTCGCGGGCTGGAG 1015
Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db 1016 AATTCATTCGTCTATATCTCTCCCTACACAGCAAGTTTCATTCTGTGCGAATGAAGGA 1075
Qy 81 AlaGlyLysSerSerTyrTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db 1076 ACTGGAAAAGTTCTTGTGGATGCTCAATCCAGAGGGTGGCAAGAGCGGGAATCTCCT 1135
Qy 100 ArgArg 101
Db 1136 AGGAGA 1141

RESULT 7
US-10-671-074-4
; Sequence 4, Application US/10671074
; Publication No. US20040097459A1
; GENERAL INFORMATION:
; APPLICANT: Dobie, Kenneth W.
; APPLICANT: Bhanot, Sanjay
; APPLICANT: Veniant-Ellison, Murielle
; APPLICANT: Lindberg, Richard A.
; APPLICANT: Shutter, John R.
; TITLE OF INVENTION: MODULATION OF FORKHEAD BOX O1A EXPRESSION
; FILE REFERENCE: AMN0001-101
; CURRENT APPLICATION NUMBER: US/10/671,074
; CURRENT FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: US 10/260,203
; PRIOR FILING DATE: 2002-09-26
; NUMBER OF SEQ ID NOS: 176
; SEQ ID NO 11
; LENGTH: 4945
; TYPE: DNA
; ORGANISM: M. musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (429)...(2387)
US-10-671-074-11

Alignment Scores:
Pred. No.: 6,79e-51 Length: 4945
Score: 426.50 Matches: 75
Percent Similarity: 91.18% Conservative: 18
```

```
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (386)...(2353)
US-10-671-074-4

Alignment Scores:
Pred. No.: 5,96e-51 Length: 5723
Score: 427.50 Matches: 75
Percent Similarity: 91.18% Conservative: 18
Best Local Similarity: 73.53% Mismatches: 8
Query Match: 77.03% Indels: 1
DB: 17 Gaps: 1

US-09-844-353A-54 (1-103) x US-10-671-074-4 (1-5723)
Qy 1 LysLysThrThrThrArgArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
Db 836 AAGAGCAGCTCGTCCCGCCGCAACGCGTGGGGCAACCTGTCTACGCGGACCTCATCACC 895
Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTrpMet 40
Db 896 AAGGCCATCGAGAGCTCGCGGAGAGCGGCTCACGCTGTGCGCAGATCTACGAGTGGATG 955
Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db 956 GTCGAAGAGCGTCCCTACTTCAAGGATAAGGGTGACAGCAACAGCTCGCGGGCTGGAG 1015
Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArgIleGlnAsnGluGly 80
Db 1016 AATTCATTCGTCTATATCTCTCCCTACACAGCAAGTTTCATTCTGTGCGAATGAAGGA 1075
Qy 81 AlaGlyLysSerSerTyrTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db 1076 ACTGGAAAAGTTCTTGTGGATGCTCAATCCAGAGGGTGGCAAGAGCGGGAATCTCCT 1135
Qy 100 ArgArg 101
Db 1136 AGGAGA 1141

RESULT 8
US-10-671-074-11
; Sequence 11, Application US/10671074
; Publication No. US20040097459A1
; GENERAL INFORMATION:
; APPLICANT: Dobie, Kenneth W.
; APPLICANT: Bhanot, Sanjay
; APPLICANT: Veniant-Ellison, Murielle
; APPLICANT: Lindberg, Richard A.
; APPLICANT: Shutter, John R.
; TITLE OF INVENTION: MODULATION OF FORKHEAD BOX O1A EXPRESSION
; FILE REFERENCE: AMN0001-101
; CURRENT APPLICATION NUMBER: US/10/671,074
; CURRENT FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: US 10/260,203
; PRIOR FILING DATE: 2002-09-26
; NUMBER OF SEQ ID NOS: 176
; SEQ ID NO 11
; LENGTH: 4945
; TYPE: DNA
; ORGANISM: M. musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (429)...(2387)
US-10-671-074-11

Alignment Scores:
Pred. No.: 6,79e-51 Length: 4945
Score: 426.50 Matches: 75
Percent Similarity: 91.18% Conservative: 18
```



```
Db 690 CGAAGGAGGCTCCCGCGAAGTCCCTGGGAATCAGTATGATGAGAACTCATCAGC 749
Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTyrMet 40
Db 750 CAGGCCATTGAAGCGCGCCGAGGAAAGCACTGACACTTGCCTCCAGATCTACGAGTGGATG 809
Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db 810 GTCCGTACTGTACCTACTTCAAGGACAAGGGTGCACCAACAGCTTCAGCAGGATGGAAG 869
Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetMetArgIleGlnAsnGluGly 80
Db 870 AACTCGATCCGCCAACCTCTTGGTGGATGCTGAACCTGAGGAGCAAGAGCGGCAAAAGCCCC 929
Qy 81 AlaGlyLysSerSerTyrTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db 930 ACCGCAAAAGCTCTTGGTGGATGCTGAACCTGAGGAGCAAGAGCGGCAAAAGCCCC 989
Qy 100 ArgArg 101
Db 990 CGCCGC 995
```

RESULT 11

```
US-10-117-722-314
; Sequence 314, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
; FILE REFERENCE: 784CIP2BCIP
; CURRENT APPLICATION NUMBER: US/10/117,722
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 314
; LENGTH: 3394
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (417)..(1934)
US-10-117-722-314
```

```
Alignment Scores:
Pred. No.: 8.6e-49 Length: 3394
Score: 410.50 Matches: 73
Percent Similarity: 86.27% Conservative: 15
Best Local Similarity: 71.57% Mismatches: 13
Query Match: 73.96% Indels: 1
DB: 16 Gaps: 1
```

US-09-844-353A-54 (1-103) x US-10-117-722-314 (1-3394)

```
Qy 1 LysLysThrThrThrArgAsnAlaTrpGlyAsnMetSerTyrAlaGluLeuIleThr 20
Db 690 CGAAGGAGGCTCCCGCGAAGTCCCTGGGAATCAGTATGATGAGAACTCATCAGC 749
Qy 21 ThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGlnValTyrGluTyrMet 40
Db 750 CAGGCCATTGAAGCGCGCCGAGGAAAGCACTGACACTTGCCTCCAGATCTACGAGTGGATG 809
```

```
Qy 41 ValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSerSerAlaGlyTrpLys 60
Db 810 GTCCGTACTGTACCTACTTCAAGGACAAGGGTGCACCAACAGCTTCAGCAGGATGGAAG 869
Qy 61 AsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetMetArgIleGlnAsnGluGly 80
Db 870 AACTCGATCCGCCAACCTCTTGGTGGATGCTGAACCTGAGGAGCAAGAGCGGCAAAAGCCCC 929
Qy 81 AlaGlyLysSerSerTyrTrpValIleAsnProAsp---AlaLysProGlyMetAsnPro 99
Db 930 ACCGCAAAAGCTCTTGGTGGATGCTGAACCTGAGGAGCAAGAGCGGCAAAAGCCCC 989
Qy 100 ArgArg 101
Db 990 CGCCGC 995
RESULT 12
US-09-205-658-44
; Sequence 44, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 2704
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-205-658-44
Alignment Scores:
Pred. No.: 2.87e-44 Length: 2704
Score: 378.50 Matches: 67
Percent Similarity: 75.93% Conservative: 15
Best Local Similarity: 62.04% Mismatches: 21
Query Match: 68.20% Indels: 5
DB: 9 Gaps: 1
US-09-844-353A-54 (1-103) x US-09-205-658-44 (1-2704)
Qy 1 LysLysThrThr-----ThrArgArgAsnAlaTrpGlyAsnMetSerTyr 15
Db 398 AAGAAGCCACCGATCAATTGCGACAGAAGAACCGAATCCATCGGGTGAGAAATCCAT 457
Qy 16 AlaGluLeuIleThrThrAlaIleMetAlaSerProGluLysArgLeuThrLeuAlaGln 35
Db 458 TCGGATATCATGTCACCAAGCATTTGGAATCGCGCCAGACGAGGCTTAAACTCAATGAG 517
Qy 36 ValTyrGluTyrMetValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSer 55
Db 518 ATTTATCAATGGTTCTCTCGATAATATTCCTACTTGTGAGAACGATCTAGTCCCGAGGAG 577
Qy 56 SerAlaGlyTyrLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArg 75
Db 578 GCCCGCGATGGAAGAACTCGATCCGTCACAAATCTGTCTCTTCATCTTCGTTTCATGCGA 637
Qy 76 IleGlnAsnGluGlyAlaGlyLysSerSerTyrTrpValIleAsnProAspAlaLysPro 95
Db 638 ATTCAGATGAAGGAGCGGAAAGAGCTCGTGGTGGTTTATTAATCCAGATGCAAGGCCA 697
Qy 96 GlyMetAsnProArgArgThrArg 103
```

```
Db 698 GGAATGAATCCACGGGTACACGT 721
|||||
RESULT 13
US-09-844-353A-44
; Sequence 44, Application US/09844353A
; Patent No. US20020037585A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351005
; CURRENT APPLICATION NUMBER: US/09/844,353A
; CURRENT FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 08/857,076
; PRIOR FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 2704
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-844-353A-44
Alignment Scores:
Pred. No.: 2,87e-44 Length: 2704
Score: 378.50 Matches: 67
Percent Similarity: 75.93% Conservative: 15
Best Local Similarity: 62.04% Mismatches: 21
Query Match: 68.20% Indels: 5
DB: 9 Gaps: 1
US-09-844-353A-54 (1-103) x US-09-844-353A-44 (1-2704)
QY 1 LysLysThrThr-----ThrArgAsnAlaTrpGlyAsnMetSerTyr 15
Db 398 AAGAAGCCACCGCATCAATTGGCAGACAGAAACCGAATCCATGGGTGAGGAATCTAT 457
QY 16 AlaGlulLeuThrThrAlaIleMetAlaSerProGlulLysArgLeuThrLeuAlaGln 35
Db 458 TCGGATATCATGCCAAAGCAATTGGAAATCGGCCAGACGGAAGGCTTAAACTCAATGAG 517
QY 36 ValTyrGlutTrpMetValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSer 55
Db 518 ATTATCAATGGTCTCTGTATATATTCCTACTTTGGAGACGATCTAGTCCCGAGGAG 577
QY 56 SerAlaGlyTrpLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArg 75
Db 578 GCGCGCGGATGGAAGAACTCGATCCGTCACAATCTGCTCTTCATCTCGTTTCATGCGA 637
QY 76 IleGlnAsnGluGlyAlaGlyLysSerSerTrpValIleAsnProAspAlaLysPro 95
Db 638 ATTCAGAAATGAAGACCGGAAAGAGCTCGTGTGGGTATTATTAATCCAGATGCAAGCCA 697
QY 96 GlyMetAsnProArgArgThrArg 103
Db 698 GGAATGAATCCACGGGTACACGT 721
|||||
RESULT 14
US-09-844-353A-44
; Sequence 44, Application US/09963693
; Patent No. US20030181364A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
```

```
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/963,693
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/205,658
; PRIOR FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 08/857,076
; PRIOR FILING DATE: 1997-05-15
; PRIOR APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 2704
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-963-693-44
Alignment Scores:
Pred. No.: 2,87e-44 Length: 2704
Score: 378.50 Matches: 67
Percent Similarity: 75.93% Conservative: 15
Best Local Similarity: 62.04% Mismatches: 21
Query Match: 68.20% Indels: 5
DB: 10 Gaps: 1
US-09-844-353A-54 (1-103) x US-09-963-693-44 (1-2704)
QY 1 LysLysThrThr-----ThrArgAsnAlaTrpGlyAsnMetSerTyr 15
Db 398 AAGAAGCCACCGCATCAATTGGCAGACAGAAACCGAATCCATGGGTGAGGAATCTAT 457
QY 16 AlaGlulLeuThrThrAlaIleMetAlaSerProGlulLysArgLeuThrLeuAlaGln 35
Db 458 TCGGATATCATGCCAAAGCAATTGGAAATCGGCCAGACGGAAGGCTTAAACTCAATGAG 517
QY 36 ValTyrGlutTrpMetValGlnAsnValProTyrPheArgAspLysGlyAspSerAsnSer 55
Db 518 ATTATCAATGGTCTCTGTATATATTCCTACTTTGGAGACGATCTAGTCCCGAGGAG 577
QY 56 SerAlaGlyTrpLysAsnSerIleArgHisAsnLeuSerLeuHisSerArgPheMetArg 75
Db 578 GCGCGCGGATGGAAGAACTCGATCCGTCACAATCTGCTCTTCATCTCGTTTCATGCGA 637
QY 76 IleGlnAsnGluGlyAlaGlyLysSerSerTrpValIleAsnProAspAlaLysPro 95
Db 638 ATTCAGAAATGAAGACCGGAAAGAGCTCGTGTGGGTATTATTAATCCAGATGCAAGCCA 697
QY 96 GlyMetAsnProArgArgThrArg 103
Db 698 GGAATGAATCCACGGGTACACGT 721
|||||
RESULT 15
US-10-186-839-1
; Sequence 1, Application US/10186839
; Publication No. US20030096957A1
; GENERAL INFORMATION:
; APPLICANT: Bowen, Benjamin R.
; APPLICANT: Whelan, James P.
; APPLICANT: Yang, Zhenyu
; TITLE OF INVENTION: AFXZeta Transcription Factor Splice Form
; FILE REFERENCE: 4-32062A/USN
; CURRENT APPLICATION NUMBER: US/10/186,839
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: 60/302,134
; PRIOR FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1353
; TYPE: DNA
```

ORGANISM: Homo Sapien
US-10-186-839-1

Alignment Scores:
Pred. No.: 2.01e-38 Length: 1353
Score: 335.50 Matches: 59
Percent Similarity: 87.65% Conservative: 12
Best Local Similarity: 72.84% Mismatches: 9
Query Match: 60.45% Indels: 1
DB: 15 Gaps: 1

US-09-844-353A-54 (1-103) x US-10-186-839-1 (1-1353)

QY	22	Ala	Ile	Met	Ala	Ser	Pro	Glu	Lys	Arg	Leu	Thr	Leu	Ala	Gln	Val	Tyr	Glu	Trp	Met	Val	41
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	172	GCC	ATT	GAA	AGC	CGC	CGG	AGG	CGA	CGA	CTG	AC	CTT	GCC	CAG	ATT	TAC	GAT	TGG	TC	231	
QY	42	Gln	Asn	Val	Pro	Tyr	Phe	Arg	Asp	Lys	Gly	Asp	Ser	Asn	Ser	Ala	Gly	Trp	Lys	Asn	61	
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	232	CGT	ACT	GTA	CC	TACT	TCA	AGG	ACA	GGG	TGC	ACA	CA	AGC	TCA	G	CAG	GAT	TGG	AAG	291	
QY	62	Ser	Ile	Arg	His	Asn	Leu	Ser	Leu	His	Ser	Arg	Phe	Met	Arg	Ile	Gln	Asn	Glu	Gly	Ala	81
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	292	TCG	ATC	CGC	CC	ACA	CTC	CTC	CTG	CA	CAG	CA	AGT	TCA	TCA	AGG	TTCA	CA	ACG	AGC	CC	351
QY	82	Gly	Lys	Ser	Ser	Trp	Trp	Val	Ile	Asn	Pro	Asp	---	Ala	Lys	Pro	Gly	Met	Asn	Pro	Arg	100
			:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	352	GGC	AAA	AGC	CTC	TTG	TGG	ATG	CTG	TAA	CCCT	GAG	GAG	GCA	AG	CAG	CGG	CAA	AGC	CCCC	CGC	411
QY	101	Arg	101																			
Db	412	CGC	414																			

Search completed: July 29, 2004, 13:22:45
Job time : 364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2004, 10:21:40 ; Search time 39 Seconds
(without alignments)
833.292 Million cell updates/sec

Title: US-09-844-353A-54

Perfect score: 555

Sequence: 1 KKTTRNAGNMSYAEILT.....SSWWINPDKPGNNPRTR 103

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phase:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriaph:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match	Length	DB ID	Description
1	549	98.9	407	5	Q8MX10	Q8mx10 caenorhabdi
2	549	98.9	508	5	O16849	O16849 caenorhabdi
3	549	98.9	510	5	O18676	O18676 caenorhabdi
4	450	81.1	613	5	Q95V55	Q95v55 drosophila
5	440.5	79.4	672	11	Q9WVH4	Q9wvh4 mus musculu
6	435.5	78.5	727	13	Q9W7F9	Q9w7f9 gallus gall
7	433.5	78.1	664	13	Q90YK2	Q90yk2 xiphophorus
8	427.5	77.0	653	11	Q810W5	Q810w5 spermophilu
9	427.5	77.0	662	6	Q8MK69	Q8mk69 sus scrofa
10	426.5	76.8	652	11	Q9JWJ4	Q9jwj4 mus musculu
11	426.5	76.8	652	11	Q9WVH5	Q9wvh5 mus musculu
12	425.5	76.7	651	13	Q9W7F8	Q9w7f8 brachydanio
13	372.5	67.1	530	5	O16850	O16850 caenorhabdi
14	339.5	61.2	333	11	Q9ES26	Q9es26 rattus norv
15	286	51.5	204	13	Q9VPK6	Q9vpk6 brachydanio
16	267.5	48.2	593	5	Q8T5L2	Q8t5l2 anopheles g

17	265	47.7	448	5	Q9VFN8	Q9vfn8 drosophila
18	221.5	39.9	623	11	Q8BUR3	Q8bur3 mus musculu
19	209.5	37.7	230	5	O17593	O17593 caenorhabdi
20	195.5	35.2	270	5	O22510	O22510 caenorhabdi
21	195	35.1	740	5	Q9VTF7	Q9vtf7 drosophila
22	193	34.8	318	5	Q8ITI5	Q8itis mmemlopsis
23	193	34.8	381	13	Q8JUT5	Q8jit5 xenopus lae
24	191	34.4	310	5	Q9TZK1	Q9tzk1 caenorhabdi
25	190	34.2	356	5	Q7YTB3	Q7ytb3 saccoglossu
26	190	34.2	373	13	Q7ZYQ0	Q7zyq0 xenopus lae
27	190	34.2	383	13	Q7TIC0	Q7tic0 brachydanio
28	188.5	34.0	271	5	Q86LT7	Q86lt7 branchiosto
29	188.5	34.0	395	15	Q7SX57	Q7sxs7 avian sarco
30	188.5	34.0	402	5	O61733	O61733 branchiosto
31	188.5	34.0	420	13	Q73862	Q73862 brachydanio
32	188.5	34.0	436	13	Q9VHC5	Q9vhc5 xenopus lae
33	188.5	34.0	478	11	Q80VP3	Q80vp3 mus musculu
34	187.5	33.8	187	13	Q800K3	Q800k3 oryzias lat
35	187.5	33.8	489	4	Q86XT7	Q86xt7 homo sapien
36	186.5	33.6	476	13	Q9DE25	Q9de25 brachydanio
37	186	33.5	345	11	Q922I5	Q922i5 mus musculu
38	186	33.5	372	11	Q9D299	Q9d299 mus musculu
39	186	33.5	451	5	Q8SZ95	Q8sz95 drosophila
40	184.5	33.2	353	13	Q73785	Q73785 brachydanio
41	184.5	33.2	492	13	Q9YHB2	Q9yhb2 xenopus lae
42	184.5	33.2	492	13	Q9PVZ3	Q9pvz3 xenopus lae
43	184.5	33.2	528	13	Q93440	Q93440 gallus gall
44	183.5	33.1	623	13	Q7T2G3	Q7t2g3 brachydanio
45	183	33.0	329	11	Q8BIK9	Q8bik9 mus musculu

ALIGNMENTS

RESULT 1

Q8MX10	PRELIMINARY;	PRT;	407 AA.
ID	Q8MX10		
AC	Q8MX10;		
DT	01-OCT-2002 (TrEMBLrel. 22, Created)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)		
DE	DAF-16 protein (corresponding sequence R13H8.1d).		
DE	DAF-16 protein (corresponding sequence R13H8.1d).		
GN	R13H8.1 OR DAF-16.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Bristol N2.		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	Waterston R.;		
RT	"Genome sequence of the nematode C. elegans: a platform for		
RT	investigating biology. The C. elegans Sequencing Consortium.;"		
RL	Science 282:2012-2018(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Bristol N2;		
RA	Jones K., Hinds K., Sutterer C., Cofman M.;		
RT	"The sequence of C. elegans cosmid R13H8.1;"		
RT	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Bristol N2;		
RA	Waterston R.;		
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF039717; AAWS1498.1; -		
DR	WormPep; R13H8.1d; CE31056.		
DR	GO; GO:0005634; C:nucleus; IEA.		
DR	GO; GO:0003700; F:transcription factor activity; IEA.		
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.		
DR	InterPro; IPR001766; TF_Pork_head.		
DR	Pfam; PF00250; Fork_head; 1.		

DR SQ

```
AC Q9W7F9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Forkhead protein FKHR.
GN FKHR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21251166; PubMed=11353388;
RA Biggs W.H. III, Cavene W.K., Arden K.C.;
RT "Identification and characterization of members of the FKHR (FOX O)
RT subclass of winged-helix transcription factors in the mouse.";
RL Mamm. Genome 12:416-425(2001).
DR EMBL; AF114261; AAD42109.1; -.
DR HSP; Q63245; 2HFH.
DR TRANSFAC; T04202; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00658; FORK HEAD 2; 1.
DR PROSITE; PS50039; FORK HEAD 3; 1.
SQ SEQUENCE 727 AA; 77290 MW; 256A2220410B075E CRC64;

Query Match 78.5%; Score 435.5; DB 13; Length 727;
Best Local Similarity 74.5%; Pred. No. 1.2e-40;
Matches 76; Conservative 18; Mismatches 7; Indels 1; Gaps 1;

Qy 1 KKTTRNANWNGMSYAEILITTAIMASPEKRLTLTAQVYEMVQNVYPRDKGDSNSSAGWK 60
Db 221 KSSSRNANWNLGYADLITKAIESSPEKRLTLTSQIYDWMVRSIPYFKDKGDSNSSAGWK 280

Qy 61 NSIRHNLSLHSRFRMRIONEAGKSGSWVINPD-AKPGMNP 101
Db 281 NSIRHNLSLHSKFRIRVQNEGTGKSSWWMLNPEGGKSGKSPRR 322

RESULT 7
Q90YK2
ID Q90YK2 PRELIMINARY; PRT; 664 AA.
AC Q90YK2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Forkhead protein xFKHR1.
GN FOXO5.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Jpl63 B85 S(A);
RA Rudd M.D., Kazianis S., Butler A.P.;
RT "Cloning and Analysis of an FKHR Transcription Factor Family Member
RT From a Xiphophorus Melanoma Model.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY040320; AAK74186.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
```

```
DR Pfam; PF00250; Fork head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00658; FORK HEAD 2; 1.
DR PROSITE; PS50039; FORK HEAD 3; 1.
SQ SEQUENCE 664 AA; 70507 MW; 268B3F3FF64D7CC0 CRC64;

Query Match 78.1%; Score 433.5; DB 13; Length 664;
Best Local Similarity 74.5%; Pred. No. 1.8e-40;
Matches 76; Conservative 17; Mismatches 8; Indels 1; Gaps 1;

Qy 1 KKTTRNANWNGMSYAEILITTAIMASPEKRLTLTAQVYEMVQNVYPRDKGDSNSSAGWK 60
Db 127 RKASSRNANWNLGYADLITKAIESSPEKRLTLTSQIYDWMVRSIPYFKDKGDSNSSAGWK 186

Qy 61 NSIRHNLSLHSRFRMRIONEAGKSGSWVINPD-AKPGMNP 101
Db 187 NSIRHNLSLHSRFRIRVQNEGTGKSSWWMLNPEGGKSGKAPRR 228

RESULT 8
Q810W5
ID Q810W5 PRELIMINARY; PRT; 653 AA.
AC Q810W5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Forkhead box O1a protein.
GN FOXO1A.
OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC Spermophilus.
OX NCBI_TaxID=43179;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Cai D., Hallenbeck J.M.;
RT "Cloning and characterization of a forkhead transcriptional factor
RT from thirteen-lined ground squirrel.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY255525; AAO72710.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00658; FORK HEAD 2; 1.
DR PROSITE; PS50039; FORK HEAD 3; 1.
SQ SEQUENCE 653 AA; 69444 MW; 31B6A6785F4F7E59 CRC64;

Query Match 77.0%; Score 427.5; DB 11; Length 653;
Best Local Similarity 73.5%; Pred. No. 8.5e-40;
Matches 75; Conservative 18; Mismatches 8; Indels 1; Gaps 1;

Qy 1 KKTTRNANWNGMSYAEILITTAIMASPEKRLTLTAQVYEMVQNVYPRDKGDSNSSAGWK 60
Db 149 KSSSRNANWNLGYADLITKAIESSAEKRLTLTSQIYEMVKSYPYFKDKGDSNSSAGWK 208

Qy 61 NSIRHNLSLHSRFRMRIONEAGKSGSWVINPD-AKPGMNP 101
Db 209 NSIRHNLSLHSKFRIRVQNEGTGKSSWWMLNPEGGKSGKSPRR 250

RESULT 9
Q8MK69
ID Q8MK69 PRELIMINARY; PRT; 662 AA.
AC Q8MK69;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
```

```
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Forkhead/winged helix transcription factor FOXO1A.
GN FOXO1A.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhu Q., Cunningham M.A., Hammond J.M.;
RT "FKHR Expression in Porcine Granulosa Cells.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY094061; AAM19156.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
SQ SEQUENCE 662 AA; 69861 MW; 96A8B841F55E5E9B CRC64;

Query Match 77.0%; Score 427.5; DB 6; Length 662;
Best Local Similarity 73.5%; Pred. No. 8.7e-40;
Matches 75; Conservative 18; Mismatches 8; Indels 1; Gaps 1;

QY 1 KKTTRRNAGNMSYAEILTTAIMASPEKRLTAQVYEMWQVYFRDKGDSNSAGWK 60
Db 158 KSSSRNAGNLSYADLTITKAIESAEKRLTUSQIYEMWVKSVPYFKDKGDSNSAGWK 217

QY 61 NSIRHNLHSRPMRIQNEGAGKSSWWVINPD-AKPGMPPR 101
Db 218 NSIRHNLHSKFIHQVNEGKSSWWMLNPEGGKSGKSPRR 259

RESULT 10
Q9JW4
ID Q9JW4 PRELIMINARY; PRT; 652 AA.
AC Q9JW4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Forkhead protein FKHR.
GN FOXO1 OR FKHR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Leenders H., Benoist C., Mathis D.;
RT "The forkhead FKHR is involved in thymocyte proliferation.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; A025157; CAB86873.1; -.
DR HSSP; Q63245; 2HFH.
DR MGD; MGI:1890077; Foxo1.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
SQ SEQUENCE 652 AA; 69518 MW; 3FF58636EA85205F CRC64;

Query Match 76.8%; Score 426.5; DB 11; Length 652;
Best Local Similarity 73.5%; Pred. No. 1.1e-39;
Matches 75; Conservative 18; Mismatches 8; Indels 1; Gaps 1;

QY 1 KKTTRRNAGNMSYAEILTTAIMASPEKRLTAQVYEMWQVYFRDKGDSNSAGWK 60
Db 148 KTSRRNAGNLSYADLTITKAIESAEKRLTUSQIYEMWVKSVPYFKDKGDSNSAGWK 207

QY 61 NSIRHNLHSRPMRIQNEGAGKSSWWVINPD-AKPGMPPR 101
Db 208 NSIRHNLHSKFIHQVNEGKSSWWMLNPEGGKSGKSPRR 249

RESULT 12
Q9W7F8
ID Q9W7F8 PRELIMINARY; PRT; 651 AA.
AC Q9W7F8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Forkhead protein FKHR.
GN FOXO5 OR FKHR.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
```

```

OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21251166; PubMed=11353388;
RA Biggs W.H. III, Cavenee W.K., Arden K.C.;
RT "Identification and characterization of members of the FKHR (FOX O) subclass of winged-helix transcription factors in the mouse.";
RL Mamm. Genome 12:418-425(2001).
DR EMBL; AF114262; AAD42110.1; -.
DR HSP; Q63245; 2HFH.
DR TRANSPAC; T04724; -.
DR ZFIN; ZDB-GENE-990708-6; foxo5.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
SQ SEQUENCE 651 AA; 69861 MW; 02A0C98BB5029C98 CRC64;

Query Match 76.7%; Score 425.5; DB 13; Length 651;
Best Local Similarity 72.5%; Pred. No. 1.4e-39;
Matches 74; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

Qy 1 KKTTRNAGNMSYAEILITTAIMASPEKRLTLAQVYEMVQNVYFRDKGDSNSAGWK 60
Db 118 RKSSRRNAGNLSYADLIITKAISTPDKRLTLQIYDMVSSVYFPAKDKGDSNSAGWK 177

Qy 61 NSIRHNLSLHSRFRMRQNEGAGKSSWWINPD-AKPGNPR 101
Db 178 NSIRHNLSLHSRFRMRQNEGAGKSSWWINPDAGRGKGKAPRR 219

RESULT 13
OL6850
ID OL6850 PRELIMINARY; PRT; 530 AA.
AC OL6850;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Fork head-related transcription factor DAF-16b (Hypothetical protein R13H8.1a).
GN DAF-16 OR R13H8.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RC MEDLINE=98013175; PubMed=9353126;
RA Ogg S., Paradis S., Gottlieb S., Patterson G.I., Lee L., Tissenbaum H.A., Ruvkun G.;
RT "The fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in C. elegans.";
RL Nature 389:994-999(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RC MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Jones K., Hinds, k, Sutterer C., Cofman M.;

```

```

RT "The sequence of C. elegans cosmid R13H8.";
OX NCBI_TaxID=7955;
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020344; AAB84392.1; -.
DR EMBL; AF039717; AAK82917.1; -.
DR PIR; T42234; T42234.
DR HSP; Q63245; 2HFH.
DR TRANSPAC; T03400; -.
DR WormPep; R13H8.1a; CE28771.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
SQ SEQUENCE 530 AA; 57898 MW; 8665FBD8428039D6 CRC64;

Query Match 67.1%; Score 372.5; DB 5; Length 530;
Best Local Similarity 61.1%; Pred. No. 1.2e-33;
Matches 66; Conservative 15; Mismatches 22; Indels 5; Gaps 1;

Qy 1 KKT-----TRNAGNMSYAEILITTAIMASPEKRLTLAQVYEMVQNVYFRDKGDSNS 55
Db 150 KKPTDQLAQKPNPGEESYDIIAKALESAPDGLKLNIEYQWFSDNIPYFGSRSPPE 209

Qy 56 SAGWKSIRHNLSLHSRFRMRQNEGAGKSSWWINPDAGKGNPRTR 103
Db 210 AAGWKSIRHNLSLHSRFRMRQNEGAGKSSWWINPDAGKGNPRTR 257

RESULT 14
Q9ES26
ID Q9ES26 PRELIMINARY; PRT; 333 AA.
AC Q9ES26;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Transcription factor FKHR (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20549654; PubMed=10960473;
RA Schmol D., Walker K.S., Alessi D.R., Grempler R., Burchell A., Guo S., Walther R., Unterman T.G.;
RT "Regulation of glucose-6-phosphatase gene expression by protein kinase Balpha and the forkhead transcription factor FKHR. Evidence for insulin response unit-dependent and -independent effects of insulin on promoter activity.";
RL J. Biol. Chem. 275:36324-36333(2000).
DR EMBL; AF247812; AAG09779.1; -.
DR HSP; Q63245; 2HFH.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.

```

```

FT NON TER 1 1
FT NON_TER 333 333
SQ SEQUENCE 333 AA; 35994 MW; D86C818F0538361D CRC64;

Query Match
Best Local Similarity 61.2%; Score 339.5; DB 11; Length 333;
Matches 59; Conservative 13; Mismatches 5; Indels 1; Gaps 1;

QY 25 ASPEKRLTLAQVYEMVQNVYFRDKGDSNSAGWKNSIRHNLHLHRPMRIQNEGAGKS 84
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 SSAEKRLLTUSQIYEMVVKSVFYFKDKGDSNSAGWKNSIRHNLHLHSKEIRVQNEGTKS 60

QY 85 SWWINPD-AKPGMPPR 101
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 SWMLNPEGKSGKSPRR 78

RESULT 15
Q9PVK6 PRELIMINARY; PRT; 204 AA.
AC Q9PVK6;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DE Putative fork head domain protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Icard-Liepkals C., Haire R.N., Strong S.J., Litman G.W.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF064828; AAD53002.1; -.
DR HSSP; Q63245; 2HFH.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; E:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
SQ SEQUENCE 204 AA; 22222 MW; BA78A856937B8CD9 CRC64;

Query Match 51.5%; Score 286; DB 13; Length 204;
Best Local Similarity 72.1%; Pred. No. 2.3e-24;
Matches 49; Conservative 15; Mismatches 4; Indels 0; Gaps 0;

QY 1 KKTTRRNAGNMVSAELITTAIMASPEKRLTLAQVYEMVQNVYFRDKGDSNSAGWK 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
118 RKSSRRNAGNGLSYADLTITKAISTPDKRLTLSQIDYDMVSSVYFYFKDKGTNSSAGWK 177

QY 61 NSIRHNL 68
Db |||||
178 YSIRHNL 185

```

Search completed: July 23, 2004, 10:25:28
Job time : 41 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2004, 10:20:10 ; Search time 54 Seconds
(without alignments)
538.933 Million cell updates/sec

Title: US-09-844-353A-54

Perfect score: 555

Sequence: 1 KKTTRRNAGNMSYAEILIT.....SSWWVINPDKPGNPRTR 103

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	555	100.0	103	3 AAB06073	Caenorhab
2	555	100.0	106	3 AAB06065	Caenorhab
3	555	100.0	109	3 AAB06075	Caenorhab
4	555	100.0	510	3 AAB06070	Caenorhab
5	517	93.2	97	4 AAB74317	C. elegans
6	440.5	79.4	673	3 AAY96449	Forkhead
7	427.5	77.0	655	3 AAY96448	Forkhead
8	427.5	77.0	655	3 AAB06076	Human hom
9	427.5	77.0	656	7 ADP47623	Human Pro
10	427.5	77.0	656	7 ADP48792	Human Pro
11	422.5	76.1	593	6 AB014656	Novel hum
12	410.5	74.0	358	4 AAU00214	Fusion pr
13	410.5	74.0	505	4 AAM39277	Human pol
14	410.5	74.0	513	4 AAM41063	Human pol
15	406.5	73.2	501	3 AAY96447	Forkhead
16	406.5	73.2	501	5 ABG95042	Human tra
17	406.5	73.2	501	5 ABG95043	Human tra
18	406.5	73.2	505	6 ABP72186	Human tra
19	378.5	68.2	509	3 AAB06071	Caenorhab
20	378	68.1	98	3 AAB06077	Caenorhab
21	367.5	66.2	99	2 AAY55739	FKHR fork
22	265	47.7	448	4 ABB59659	Drosophil
23	222.5	40.1	588	6 AAE37024	Human nuc
24	195	35.1	740	4 ABB66642	Drosophil
25	195	35.1	740	4 ABB66641	Drosophil

RESULT 1

AAB06073

ID AAB06073 standard; protein; 103 AA.

XX AC AAB06073;

XX DT 11-JAN-2001 (first entry)

XX DE Caenorhabditis elegans DAF-16 conserved motif #1.

XX KW Caenorhabditis elegans; daf-16; daf-18; insulin signalling pathway;
KW daf-2; age-1; insulin receptor; PI 3-kinase; PKB kinase; AKT kinase;
KW PTEN lipid phosphatase; antidiabetic; anorectic; obesity; diabetes;
KW impaired glucose tolerance; transgenic animal; conserved motif.

XX OS Caenorhabditis elegans.

XX PN WC2000033068-A1.

XX PD 08-JUN-2000.

XX PF 02-DEC-1999; 99WO-US028529.

XX PR 03-DEC-1998; 98US-00205658.

XX PA (GEO) GEN HOSPITAL CORP.

XX PI Ruvkun G, Ogg S;

XX WPI; 2000-423022/36.

XX PT Diagnosing and treating obesity and impaired glucose tolerance using
modulators of daf-18 expression and/or activity.

XX PS Disclosure; Page 134; 402pp; English.

XX CC The present sequence is a conserved motif in DAF-16 from Caenorhabditis
elegans. DAF-16 is the major transcriptional output of the C. elegans
insulin signalling pathway. A number of C. elegans genes have been
identified as homologues of genes in the mammalian insulin signalling
pathway. The C. elegans age-1 gene encodes a homologue of the mammalian
PI 3-kinase whilst daf-2 encodes a homologue of the mammalian insulin
receptor. The C. elegans PKB kinase and AKT kinase act downstream of daf-
2 and age-1, just as their mammalian homologues act downstream of insulin
signalling. Other daf genes have also been implicated in the C. elegans
insulin signalling pathway. The C. elegans PTEN lipid phosphatase
homologue, DAF-18, has been found to act upstream of AKT in the pathway.
This discovery has enabled mammalian PTEN action to be mapped to the

ALIGNMENTS

26	195	35.1	740	4	ABB65910	Abb65910 Drosophil
27	190.5	34.3	106	2	AAY55734	Aay55734 FKHL4 for
28	189.5	34.1	190	4	AAU16159	Aau16159 Human nov
29	189.5	34.1	190	4	AAU16577	Aau16577 Human nov
30	189.5	34.1	190	4	ABB15369	Abb15369 Human nov
31	189.5	34.1	190	6	ABU55228	Abu55228 Human nov
32	189.5	34.1	190	6	ABU55646	Abu55646 Human nov
33	188.5	34.0	477	6	ABP72344	Abp72344 Brain fac
34	188.5	34.0	480	2	AAR44551	Aar44551 Brain fac
35	188.5	34.0	481	5	ABB57076	Abb57076 Mouse isc
36	187.5	33.8	105	2	AAY55735	Aay55735 FKHL2 for
37	187.5	33.8	365	4	ABB57872	Abb57872 Drosophil
38	186	33.5	451	4	ABB59467	Abb59467 Drosophil
39	184.5	33.2	106	2	AAY55731	Aay55731 FKHL17 fo
40	184.5	33.2	106	2	AAY55730	Aay55730 FKHL8 for
41	184.5	33.2	106	2	AAY55723	Aay55723 FKHL7 for
42	183.5	33.1	106	2	AAY55724	Aay55724 FKHL14 fo
43	183.5	33.1	431	7	ADE09208	Ade09208 Novel pro
44	182.5	32.9	465	7	ADB75312	Adb75312 Prostata
45	182.5	32.9	553	2	AAY43260	Aay43260 Human FKH

QY 1 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEWVQVVPYFRDKGDSNSAGWK 60
DB 7 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEWVQVVPYFRDKGDSNSAGWK 66
QY 61 NSIRHNLSLHSRPMRIQNEGAGKSSWWVINPDAPGKNPRTTR 103
DB 67 NSIRHNLSLHSRPMRIQNEGAGKSSWWVINPDAPGKNPRTTR 109
RESULT 4
ID RAB06070 standard; protein; 510 AA.
XX AC RAB06070;
XX
DT 11-JAN-2001 (first entry)
DE XX
DE XX Caenorhabditis elegans DAF-16 isoform #1.
KW Caenorhabditis elegans; daf-16; daf-2; age-1; daf-18;
KW insulin signalling pathway; insulin receptor; PI 3-kinase; PKB kinase;
KW AKT kinase; PTEN lipid phosphatase; antidiabetic; anorectic; obesity;
KW diabetes; impaired glucose tolerance; transgenic animal.
XX
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT Misc-difference 119 /note= "encoded by CTT"
FT Misc-difference 120 /note= "encoded by CGG"
FT Misc-difference 121 /note= "encoded by AGT"
FT Misc-difference 295 /note= "encoded by GA"
XX
XX WO200033068-A1.
XX
XX
PD 08-JUN-2000.
XX
XX
PF 02-DEC-1999; 99WO-US028529.
XX
XX
PR 03-DEC-1998; 98US-00205658.
XX
XX (GEO) GEN HOSPITAL CORP.
XX
XX Ruvkun G, Ogg S;
XX
XX WPI; 2000-423022/36.
DR N-PSDB; AAA62298.
XX
XX
PT Diagnosing and treating obesity and impaired glucose tolerance using
PT modulators of daf-18 expression and/or activity.
XX
XX
PS Disclosure; Fig 14A; 402pp; English.
XX
XX The present cDNA sequence from Caenorhabditis elegans corresponds to a
XX differentially spliced daf-16 transcript which encodes an isoform of DAF-
XX 16. DAF-16 interacts with the C. elegans insulin signalling pathway. A
XX number of C. elegans genes have been identified as homologues of genes in
XX the mammalian insulin signalling pathway. The C. elegans age-1 gene
XX encodes a homologue of the mammalian PI 3-kinase whilst daf-2 encodes a
XX homologue of the mammalian insulin receptor. The C. elegans PKB kinase
XX and AKT kinase act downstream of daf-2 and age-1, just as their mammalian
XX homologues act downstream of insulin signalling. Other daf genes have
XX also been implicated in the C. elegans insulin signalling pathway. The C.
XX elegans PTEN lipid phosphatase homologue, DAF-18, has been found to act
XX upstream of AKT in the pathway. This discovery has enabled mammalian PTEN
XX action to be mapped to the insulin signalling pathway. Compounds that
XX inhibit the expression and/or activity of polypeptides encoded by these
XX genes may be administered to patients to treat or prevent disorders such
XX as obesity and impaired glucose tolerance

XX SQ Sequence 510 AA;
Query Match 100.0%; Score 555; DB 3; Length 510;
Best Local Similarity 100.0%; Pred. No. 6.2e-59;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEWVQVVPYFRDKGDSNSAGWK 60
DB 135 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEWVQVVPYFRDKGDSNSAGWK 194
QY 61 NSIRHNLSLHSRPMRIQNEGAGKSSWWVINPDAPGKNPRTTR 103
DB 195 NSIRHNLSLHSRPMRIQNEGAGKSSWWVINPDAPGKNPRTTR 237
RESULT 5
RAB74317
ID AAB74317 standard; peptide; 97 AA.
XX AC AAB74317;
XX
DT 27-JUN-2001 (first entry)
XX
DE C. elegans DAF-16 winged helix domain.
XX
KW DAF-16; winged helix domain; homolog; ageing; ischemia;
KW cardiac hypertrophy; fever; inflammation; metabolic disease; viral;
KW bacterial; infection; cancer; tumour; stress.
OS Caenorhabditis elegans.
XX
XX WO200118549-A1.
XX
XX PD 15-MAR-2001.
XX
XX PF 07-SEP-2000; 2000WO-US024487.
XX
XX PR 07-SEP-1999; 99US-0152825P.
XX
XX (NEUR-) NEUROGENETICS INC.
XX
XX PI Henderson ST, Johnson TE;
XX
XX WPI; 2001-226755/23.
XX
XX Identifying agents that can increase the activity of DAF-16 or their
XX mammalian homologs, useful for inhibiting tumor growth and for protecting
XX the cells of an organism from environmental stress, e.g. heat and
XX cytotoxic drugs.
XX
XX Disclosure; Page 10; 59pp; English.
XX
XX The present invention relates to methods of identifying agents that can
XX increase the activity of C.elegans DAF-16 or their mammalian homologs.
XX Compounds identified by the methods are useful in any disease state where
XX tissue damage results or in ageing, for inhibiting certain tumour growth
XX and for protecting the cells of an organism from damage associated with
XX diseases, e.g. ischemia, cardiac hypertrophy, fever, inflammation,
XX metabolic diseases, viral and bacterial infection, cell and tissue
XX trauma, and cancer. The compounds are also useful for protecting the
XX cells of an organism from environmental stress. The present sequence is
XX the C. elegans DAF-16 winged helix domain. This peptide contains a
XX putative DNA binding region
XX
XX SQ Sequence 97 AA;
Query Match 93.2%; Score 517; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 3.1e-55;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEWVQVVPYFRDKGDSNSAGWK 60

Db 1 KKTTRRNAGNMSYAEILITTAIMASPEKRLTLAQVYEMWQNVYFRDKGDSNSAGWK 60
 QY 61 NSIRHNLHLRFRMRQNEGAGKSSWWVINPDAPG 96
 DB 61 NSIRHNLHLRFRMRQNEGAGKSSWWVINPDAPG 96

RESULT 6
 AAY96449
 ID AAY96449 standard; protein; 673 AA.
 XX
 AC AAY96449;
 XX
 DT 12-SEP-2000 (first entry)
 XX
 DE Forkhead transcription factor FKHL1.
 XX
 KW Protein kinase B; PKB; substrate; phosphorylation; forkhead; Afx; FKHL1;
 KW transcription factor; insulin; suppressor.
 XX
 OS Homo sapiens.
 XX
 PN WO200031291-A1.
 XX
 PD 02-JUN-2000.
 XX
 PF 16-NOV-1999; 99WO-SE002095.
 XX
 PR 19-NOV-1998; 98SE-00003958.
 XX
 PA (PHAA) PHARMACIA & UPJOHN AB.
 XX
 PI James S, Dahlman-Wright K, Lake S, Butcher S, Climent I;
 XX
 DR WPI; 2000-400095/34.
 XX
 PT Method for screening substances which modulate Protein kinase B activity
 PT uses substrate peptides.
 XX
 PS Disclosure; Page 19-22; 41pp; English.
 XX
 CC This is the forkhead transcription factor, FKHL1, which is related to
 CC Caenorhabditis elegans nematode worm forkhead protein, daf-16. FKHL1 is
 CC related to Afx (AAY96447), which is phosphorylated by PKB. This is
 CC enhanced when cells are stimulated with insulin or IGF-1. PKB therefore
 CC demonstrates a novel regulation of transcription. Afx-mediated
 CC stimulation of IRE driven reporter genes is suppressed by overexpression
 CC of PKB and this is further suppressed by insulin. This mechanism can be
 CC used to identify compounds for treatment of metabolic disease. Substrate
 CC peptides (AAY96441-46) derived from Afx, and forkhead proteins FKHL
 CC (AAY96448) and FKHL1 can be used to discriminate between the effects of
 CC compounds which mediate insulin action through transcription via forkhead
 CC transcription factor family from those which modulate activity of enzymes
 CC involved in metabolism by phosphorylation. The claimed method comprises
 CC screening for substances which are activators, inhibitors or binders of
 CC PKB by the use of 2 sequences chosen from AAY96441 and AAY96442, AAY96443
 CC and AAY96444 and AAY96445 and AAY96446

Query Match 79.4%; Score 440.5; DB 3; Length 673;
 Best Local Similarity 77.5%; Pred. No. 9.7e-45;
 Matches 79; Conservative 13; Mismatches 9; Indels 1; Gaps 1;
 QY 1 KKTTRRNAGNMSYAEILITTAIMASPEKRLTLAQVYEMWQNVYFRDKGDSNSAGWK 60
 DB 148 KCSRRNAGNLSYADLITRAIESPDKRLTLAQVYEMWQNVYFRDKGDSNSAGWK 207

RESULT 8
 AAB06076
 ID AAB06076 standard; protein; 655 AA.
 XX
 AC AAB06076;

RESULT 7
 AAY96448
 ID AAY96448 standard; protein; 655 AA.
 XX
 AC AAY96448;
 XX
 DT 12-SEP-2000 (first entry)
 XX
 DE Forkhead transcription factor FKHL.
 XX
 KW Protein kinase B; PKB; substrate; phosphorylation; forkhead; Afx; FKHL;
 KW transcription factor; insulin; suppressor.
 XX
 OS Homo sapiens.
 XX
 PN WO200031291-A1.
 XX
 PD 02-JUN-2000.
 XX
 PF 16-NOV-1999; 99WO-SE002095.
 XX
 PR 19-NOV-1998; 98SE-00003958.
 XX
 PA (PHAA) PHARMACIA & UPJOHN AB.
 XX
 PI James S, Dahlman-Wright K, Lake S, Butcher S, Climent I;
 XX
 DR WPI; 2000-400095/34.
 XX
 PT Method for screening substances which modulate Protein kinase B activity
 PT uses substrate peptides.
 XX
 PS Disclosure; Page 16-19; 41pp; English.
 XX
 CC This is the forkhead transcription factor, FKHL, which is related to the
 CC Caenorhabditis elegans nematode worm forkhead protein, daf-16. FKHL is
 CC phosphorylated by protein kinase B (PKB). FKHL is related to Afx
 CC (AAY96447), which is also phosphorylated by PKB. This phosphorylation, in
 CC Afx, is enhanced when cells are stimulated with insulin or IGF-1. PKB
 CC therefore demonstrates a novel regulation of transcription. Afx-
 CC mediated stimulation of IRE driven reporter genes is suppressed by
 CC overexpression of PKB and this is further suppressed by insulin. This
 CC mechanism can be used to identify compounds for treatment of metabolic
 CC disease. Substrate peptides (AAY96441-46) derived from Afx, and forkhead
 CC proteins FKHL and FKHL1 (AAY96449) can be used to discriminate between
 CC the effects of compounds which mediate insulin action through
 CC transcription via forkhead transcription factor family from those which
 CC modulate activity of enzymes involved in metabolism by phosphorylation.
 CC The claimed method comprises screening for substances which are
 CC activators, inhibitors or binders of PKB by the use of 2 sequences chosen
 CC from AAY96441 and AAY96442, AAY96443 and AAY96444 and AAY96445 and
 CC AAY96446

Query Match 77.0%; Score 427.5; DB 3; Length 655;
 Best Local Similarity 73.5%; Pred. No. 3.7e-43;
 Matches 75; Conservative 18; Mismatches 8; Indels 1; Gaps 1;
 QY 1 KKTTRRNAGNMSYAEILITTAIMASPEKRLTLAQVYEMWQNVYFRDKGDSNSAGWK 60
 DB 151 KSSRRNAGNLSYADLITRAIESAEKRLTLAQVYEMWQNVYFRDKGDSNSAGWK 210

RESULT 8
 AAB06076
 ID AAB06076 standard; protein; 655 AA.
 XX
 AC AAB06076;

CC particularly in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, which includes a pathology associated
 CC with NOVX polypeptide. The DNA encoding the protein is useful in gene
 CC therapy for treating the disease or condition. In particular, the NOVX
 CC polypeptide or polynucleotide is useful for treating endocrine/
 CC metabolism-related diseases (e.g. obesity or diabetes), central nervous
 CC system disorders (e.g. Alzheimer's disease, Parkinson's disease,
 CC epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune
 CC and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,
 CC asthma, inflammatory bowel disease, rheumatoid arthritis or
 CC osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,
 CC prostate or brain cancers, or melanoma), liver diseases (e.g. liver
 CC cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),
 CC haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
 CC These are also useful in developing powerful assay system for functional
 CC analysis of various human disorders, as well as in diagnostic
 CC applications, and for monitoring the effects of drugs during clinical
 CC trials. This is the amino acid sequence of a novel human NOV protein
 CC
 SQ Sequence 593 AA;

Query Match 76.1%; Score 422.5; DB 6; Length 593;
 Best Local Similarity 80.0%; Pred. No. 1.3e-42;
 Matches 76; Conservative 10; Mismatches 8; Indels 1; Gaps 1;
 QY 8 NAWGNMYSYAEILTTAIMASPEKRLTLAQQVYEWVQVYFDRKGDNSNSAGWKNIRHNL 67
 Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 75 NAWGNLSYADLIITRAESSFDKRLTLISQIYEWVRCVPYFDRKGDNSNSAGWKNIRHNL 134
 QY 68 SLHSRFRMRIONEGAGKSSWWVINPD-AKPGMNP 101
 Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 135 SLHSRFRMRVQNEGKSSWWIINPDGKSGKAPRR 169

RESULT 12
 AAU00214
 ID AAU00214 standard; protein; 358 AA.
 AC AAU00214;
 DT 09-MAY-2001 (first entry)
 XX Fusion protein GST-AfxDBD, amino acid sequence.
 DE
 XX DNA binding domain; Afx; transcription factor; GST-AfxDBD;
 KW human fork head transcription factor; Afx response element; human;
 KW diabetes; drug target; insulin; insulin receptor signalling pathway.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS Chimeric.

FH Key Location/Qualifiers
 FT Misc-difference 28 /note= "Encoded by TAT"
 FT Misc-difference 349 /note= "Encoded by AG"
 FT
 FN WO200114544-A1.
 EN
 PD 01-MAR-2001.
 XX
 PF 22-AUG-2000; 2000WO-SE001603.
 XX
 PR 26-AUG-1999; 99SE-00003009.
 PR 31-AUG-1999; 99US-0151867P.
 PR 25-NOV-1999; 99SE-00004269.
 XX
 PA (PHAA) PHARMACIA & UPJOHN AB.
 XX
 PI Clintment-Johanson I, Dahlman-Wright K, Lake S, Wasserman W;
 XX WPI; 2001-218446/22.
 DR

DR N-PSDB; AAS00244.

XX New Afx response element with a nucleotide sequence comprising a DNA
 PT binding site for the human fork head transcription factor Afx, useful in
 PT screening for genes or in bioinformatic analysis of the human genome.
 XX
 PS Example 2; Fig 7; 34pp; English.

XX The sequence represents the amino acid sequence of the GST-AfxDBD fusion
 CC protein encoded by expression vector pGEX-AfxDBD. Human fork head
 CC transcription factor, Afx, was expressed and the protein used to find a
 CC response element comprising an 8 base pair (bp) nucleotide sequence. The
 CC nucleotide sequence comprises AACATGTT, the Afx response element, and is
 CC useful in bioinformatic analysis e.g. of the human genome. Employing the
 CC Afx response element is also useful for screening for genes that may be
 CC used as diabetes drug targets. This can provide a subset of genes
 CC transcriptionally responsive to insulin and may lead to development of
 CC assays that facilitate the analysis of genes interacting with the insulin
 CC receptor pathway. The genes found in such screening may also be employed
 CC in further screening methods for compounds which modify the insulin
 CC receptor signalling pathway
 XX

SQ Sequence 358 AA;

Query Match 74.0%; Score 410.5; DB 4; Length 358;
 Best Local Similarity 71.6%; Pred. No. 2e-41;
 Matches 73; Conservative 15; Mismatches 13; Indels 1; Gaps 1;

QY 1 KTTTTRNAGNMYSYAEILTTAIMASPEKRLTLAQQVYEWVQVYFDRKGDNSNSAGWK 60
 Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 239 RKGGRNAGNQSAYELISQALSAPEKRLTLAQQIYEWVTRVPYFDRKGDNSNSAGWK 298
 QY 61 NSIRHNLSLHSRFRMRIONEGAGKSSWWVINPD-AKPGMNP 101
 Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 299 NSIRHNLSLHSKFIKVHNEATGKSSWWMLNPEGGKSGKAPRR 340

RESULT 13
 AAM39277
 ID AAM39277 standard; protein; 505 AA.
 AC AAM39277;
 DT 22-OCT-2001 (first entry)
 XX Human polypeptide SEQ ID NO 2422.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

OS Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

XX 21-JAN-2000; 2000US-00488725.

XX 25-APR-2000; 2000US-00552317.

XX 20-JUN-2000; 2000US-00598042.

XX 19-JUL-2000; 2000US-00620312.

XX 03-AUG-2000; 2000US-00653450.

XX 14-SEP-2000; 2000US-00662191.

XX 19-OCT-2000; 2000US-00693036.

XX 29-NOV-2000; 2000US-00727344.

PA (HYSE-) HYSEQ INC.

RESULT 14	
AAM41063	
ID	AAM41063 standard; protein; 513 AA.
XX	
AC	AAM41063;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polypeptide SEQ ID NO 5994.
XX	
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Sly-brager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX	leukaemia.
XX	
OS	Homo sapiens.
XX	
FN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US034263.
XX	
PR	23-DEC-1999; 99US-00471275.
PR	21-JAN-2000; 2000US-00488725.
PR	25-APR-2000; 2000US-00552317.
PR	20-JUN-2000; 2000US-00598042.

RESULT	15
AAY96447	
ID	AAY96447 standard; protein; 501 AA.
XX	
AC	AAY96447;
XX	
DT	12-SEP-2000 (first entry)
XX	
DE	Forkhead transcription factor Afx1-501.
XX	
KW	Protein kinase B; PKB; substrate; phosphorylation; forkhead; Afx;
XX	transcription factor; insulin; suppressor.
XX	
OS	Homo sapiens.
XX	
PN	WO200031291-A1.
XX	
PD	02-JUN-2000.
XX	
PF	16-NOV-1999; 99WO-S5002095.
XX	
PR	19-NOV-1998; 98SE-00003958.
XX	

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2004, 10:23:16 ; Search time 18 Seconds
(without alignments)
295.415 Million cell updates/sec

Title: US-09-844-353A-54

Perfect score: 555

Sequence: 1 KKTTRRNAGNMSYAEIIT.....SSWWVINPDAPGMNPRTR 103

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/6C_COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	555	100.0	103	3	US-08-857-076-54
2	555	100.0	106	3	US-08-857-076-37
3	555	100.0	109	3	US-08-857-076-56
4	555	100.0	510	3	US-08-857-076-45
5	427.5	77.0	655	3	US-08-857-076-57
6	410.5	74.0	358	4	US-09-645-623-23
7	406.5	73.2	501	3	US-08-857-076-102
8	390.5	70.4	99	3	US-09-083-351-21
9	390.5	68.2	509	3	US-08-857-076-46
10	378.5	68.2	635	3	US-08-857-076-101
11	378.5	68.1	98	3	US-08-857-076-58
12	378	68.1	106	3	US-09-083-351-16
13	188.5	34.0	106	3	US-09-083-352-16
14	188.5	34.0	106	3	US-09-083-351-17
15	185.5	33.4	106	3	US-09-083-352-17
16	185.5	33.4	480	1	US-07-882-292-2
17	185.5	33.4	480	2	US-08-331-644-2
18	185.5	33.4	480	5	PCR-US93-04102-2
19	185.5	32.9	106	3	US-09-083-351-5
20	182.5	32.9	106	3	US-09-083-351-12
21	182.5	32.9	106	3	US-09-083-351-13
22	182.5	32.9	106	3	US-09-083-352-12
23	182.5	32.9	106	3	US-09-083-352-13
24	182.5	32.9	106	3	US-09-083-352-12
25	182.5	32.9	106	3	US-09-083-351-2
26	182.5	32.9	553	3	US-09-083-351-2
27	182.5	32.9	553	3	US-09-083-352-2

```

28      182      32.8      763      2      US-08-742-753-4      Sequence 4, Appli
29      181.5      32.7      106      3      US-09-083-351-6      Sequence 6, Appli
30      181.5      32.7      106      3      US-09-083-352-6      Sequence 6, Appli
31      178.5      32.2      98      1      US-07-906-930E-4      Sequence 4, Appli
32      178.5      32.2      540      1      US-07-906-930E-2      Sequence 2, Appli
33      177.5      32.0      106      3      US-09-083-351-18      Sequence 18, Appl
34      177.5      32.0      106      3      US-09-083-352-18      Sequence 18, Appl
35      176.5      31.8      98      1      US-07-906-930E-6      Sequence 6, Appli
36      176.5      31.8      494      1      US-07-906-930E-33      Sequence 33, Appli
37      176      31.7      429      4      US-09-372-668-2      Sequence 2, Appli
38      176      31.7      429      4      US-09-697-377-2      Sequence 2, Appli
39      176      31.7      429      4      US-09-696-868-2      Sequence 2, Appli
40      176      31.7      429      4      US-09-697-341-2      Sequence 2, Appli
41      175.5      31.6      106      3      US-09-083-351-14      Sequence 14, Appl
42      175.5      31.6      106      3      US-09-083-352-14      Sequence 14, Appl
43      175.5      31.6      473      3      US-08-857-076-99      Sequence 99, Appli
44      175      31.5      431      4      US-09-372-668-4      Sequence 4, Appli
45      175      31.5      431      4      US-09-697-377-4      Sequence 4, Appli

```

ALIGNMENTS

RESULT 1
US-08-857-076-54
; Sequence 54, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:

; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Kouitarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison

; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15

; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans

US-08-857-076-54

Query Match 100.0%; Score 555; DB 3; Length 103;
Best Local Similarity 100.0%; Pred. No. 8.8e-62;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 KKTTRRNAGNMSYAEIITTTAIMASPEKRLTLAQVYEMWVONVYFRDKGDSNAGWK 60
Db      1 KKTTRRNAGNMSYAEIITTTAIMASPEKRLTLAQVYEMWVONVYFRDKGDSNAGWK 60

```

```

QY      61 NSIRHNLSHSRFMRIQNEGAGKSSWWVINPDAPGMNPRTR 103
Db      61 NSIRHNLSHSRFMRIQNEGAGKSSWWVINPDAPGMNPRTR 103

```

RESULT 2
US-08-857-076-37
; Sequence 37, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:

; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Kouitarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi

APPLICANT: Kimura, Koutarou
APPLICANT: Patterson, Garth
APPLICANT: Ogg, Scott
APPLICANT: Paradis, Suzanne
APPLICANT: Tissenbaum, Heidi
APPLICANT: Morris, Jason
APPLICANT: Koweek, Allison
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
FILE REFERENCE: 00786/351001
CURRENT APPLICATION NUMBER: US/08/857,076C
CURRENT FILING DATE: 1997-05-15
NUMBER OF SEQ ID NOS: 114
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 45
LENGTH: 510
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-08-857-076-45

Query Match 100.0%; Score 555; DB 3; Length 510;
Best Local Similarity 100.0%; Pred. No. 7.7e-61;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFRDKGDSNSSAGWK 60
Db 135 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFRDKGDSNSSAGWK 194

Qy 61 NSIRHNLHLSRFRMRIONEAGKSSWWVINPDAPGNNPRTR 103
Db 195 NSIRHNLHLSRFRMRIONEAGKSSWWVINPDAPGNNPRTR 237

RESULT 5
US-08-857-076-57
Sequence 57, Application US/08857076C
Patent No. 6225120
GENERAL INFORMATION:
APPLICANT: Ruvkun, Gary
APPLICANT: Kimura, Koutarou
APPLICANT: Patterson, Garth
APPLICANT: Ogg, Scott
APPLICANT: Paradis, Suzanne
APPLICANT: Tissenbaum, Heidi
APPLICANT: Morris, Jason
APPLICANT: Koweek, Allison
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
FILE REFERENCE: 00786/351001
CURRENT APPLICATION NUMBER: US/08/857,076C
CURRENT FILING DATE: 1997-05-15
NUMBER OF SEQ ID NOS: 114
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 57
LENGTH: 655
TYPE: PRT
ORGANISM: Homo sapiens
US-08-857-076-57

Query Match 77.0%; Score 427.5; DB 3; Length 655;
Best Local Similarity 73.5%; Pred. No. 9.2e-45;
Matches 75; Conservative 18; Mismatches 8; Indels 1; Gaps 1;

Qy 1 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFRDKGDSNSSAGWK 60
Db 151 KSSSRNAGNLSYADLTITKAIESSAEKRLTSLQIYEMVKSVPYFKDKGDSNSSAGWK 210

Qy 61 NSIRHNLHLSRFRMRIONEAGKSSWWVINPD-APGNNPRR 101
Db 211 NSIRHNLHLSKFRVQNEGTGKSSWWMLNPEGGKSGKSPRR 252

RESULT 6

APPLICANT: Morris, Jason
APPLICANT: Koweek, Allison
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
FILE REFERENCE: 00786/351001
CURRENT APPLICATION NUMBER: US/08/857,076C
CURRENT FILING DATE: 1997-05-15
NUMBER OF SEQ ID NOS: 114
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 37
LENGTH: 106
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-08-857-076-37

Query Match 100.0%; Score 555; DB 3; Length 106;
Best Local Similarity 100.0%; Pred. No. 9.1e-62;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFRDKGDSNSSAGWK 60
Db 1 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFRDKGDSNSSAGWK 60

Qy 61 NSIRHNLHLSRFRMRIONEAGKSSWWVINPDAPGNNPRTR 103
Db 61 NSIRHNLHLSRFRMRIONEAGKSSWWVINPDAPGNNPRTR 103

RESULT 3
US-08-857-076-56
Sequence 56, Application US/08857076C
Patent No. 6225120
GENERAL INFORMATION:
APPLICANT: Ruvkun, Gary
APPLICANT: Kimura, Koutarou
APPLICANT: Patterson, Garth
APPLICANT: Ogg, Scott
APPLICANT: Paradis, Suzanne
APPLICANT: Tissenbaum, Heidi
APPLICANT: Morris, Jason
APPLICANT: Koweek, Allison
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
FILE REFERENCE: 00786/351001
CURRENT APPLICATION NUMBER: US/08/857,076C
CURRENT FILING DATE: 1997-05-15
NUMBER OF SEQ ID NOS: 114
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 56
LENGTH: 109
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-08-857-076-56

Query Match 100.0%; Score 555; DB 3; Length 109;
Best Local Similarity 100.0%; Pred. No. 9.5e-62;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFRDKGDSNSSAGWK 60
Db 7 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFRDKGDSNSSAGWK 66

Qy 61 NSIRHNLHLSRFRMRIONEAGKSSWWVINPDAPGNNPRTR 103
Db 67 NSIRHNLHLSRFRMRIONEAGKSSWWVINPDAPGNNPRTR 109

RESULT 4
US-08-857-076-45
Sequence 45, Application US/08857076C
Patent No. 6225120
GENERAL INFORMATION:
APPLICANT: Ruvkun, Gary

US-09-645-629-23

Sequence 23, Application US/09645629

Patent No. 6472515

GENERAL INFORMATION:

APPLICANT: Climent-Johansson, Isabel

APPLICANT: Dahlman-Wright, Karin

APPLICANT: Lake, Staffan

APPLICANT: Wasserman, Wyeth

TITLE OF INVENTION: NOVEL RESPONSE ELEMENT

FILE REFERENCE: 13425-032001

CURRENT APPLICATION NUMBER: US/09/645,629

CURRENT FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: SE 9904269-9

PRIOR FILING DATE: 1999-11-25

PRIOR APPLICATION NUMBER: US 60/151,867

PRIOR FILING DATE: 1999-08-31

PRIOR APPLICATION NUMBER: SE 9903009-0

PRIOR FILING DATE: 1999-08-26

NUMBER OF SEQ ID NOS: 34

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 23

LENGTH: 358

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: synthetically generated GST-AfxDBD construct

US-09-645-629-23

Query Match 74.0%; Score 410.5; DB 4; Length 358;

Best Local Similarity 71.6%; Pred. No. 5.4e-43;

Matches 73; Conservative 15; Mismatches 13; Indels 1; Gaps 1;

QY 1 KKTTRNAGNMSYAEILTTAIMASPEKRLTLAQYVEMVQVYFDRKGDSSAGWK 60

DB 239 RKGGRRNAGNQSYAEILTTAIMASPEKRLTLAQYVEMVQVYFDRKGDSSAGWK 298

QY 61 NSIRHNLHSRPMRIQNEGAGKSSWWINPD-AKPGMNP 101

DB 299 NSIRHNLHSRPMRIQNEGAGKSSWWINPD-AKPGMNP 340

RESULT 7

US-08-857-076-102

Sequence 102, Application US/08857076C

Patent No. 6225120

GENERAL INFORMATION:

APPLICANT: Ruvkun, Gary

APPLICANT: Kimura, Koutarou

APPLICANT: Patterson, Garth

APPLICANT: Ogg, Scott

APPLICANT: Paradis, Suzanne

APPLICANT: Tissenbaum, Heidi

APPLICANT: Morris, Jason

APPLICANT: Kowek, Allison

TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR

TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS

FILE REFERENCE: 00786/351001

CURRENT APPLICATION NUMBER: US/08/857,076C

CURRENT FILING DATE: 1997-05-15

NUMBER OF SEQ ID NOS: 114

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 102

LENGTH: 501

TYPE: PRT

ORGANISM: Homo sapiens

US-08-857-076-102

Query Match 73.2%; Score 406.5; DB 3; Length 501;

Best Local Similarity 70.6%; Pred. No. 2.7e-42;

Matches 72; Conservative 15; Mismatches 14; Indels 1; Gaps 1;

QY 1 KKTTRNAGNMSYAEILTTAIMASPEKRLTLAQYVEMVQVYFDRKGDSSAGWK 60

DB 239 RKGGRRNAGNQSYAEILTTAIMASPEKRLTLAQYVEMVQVYFDRKGDSSAGWK 298

QY 61 NSIRHNLHSRPMRIQNEGAGKSSWWINPD-AKPGMNP 101

DB 299 NSIRHNLHSRPMRIQNEGAGKSSWWINPD-AKPGMNP 340

US-09-645-629-23

Sequence 23, Application US/09645629

Patent No. 6472515

GENERAL INFORMATION:

APPLICANT: Climent-Johansson, Isabel

APPLICANT: Dahlman-Wright, Karin

APPLICANT: Lake, Staffan

APPLICANT: Wasserman, Wyeth

TITLE OF INVENTION: NOVEL RESPONSE ELEMENT

FILE REFERENCE: 13425-032001

CURRENT APPLICATION NUMBER: US/09/645,629

CURRENT FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: SE 9904269-9

PRIOR FILING DATE: 1999-11-25

PRIOR APPLICATION NUMBER: US 60/151,867

PRIOR FILING DATE: 1999-08-31

PRIOR APPLICATION NUMBER: SE 9903009-0

PRIOR FILING DATE: 1999-08-26

NUMBER OF SEQ ID NOS: 34

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 23

LENGTH: 358

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: synthetically generated GST-AfxDBD construct

US-09-645-629-23

Query Match 74.0%; Score 410.5; DB 4; Length 358;

Best Local Similarity 71.6%; Pred. No. 5.4e-43;

Matches 73; Conservative 15; Mismatches 13; Indels 1; Gaps 1;

QY 1 KKTTRNAGNMSYAEILTTAIMASPEKRLTLAQYVEMVQVYFDRKGDSSAGWK 60

DB 239 RKGGRRNAGNQSYAEILTTAIMASPEKRLTLAQYVEMVQVYFDRKGDSSAGWK 298

QY 61 NSIRHNLHSRPMRIQNEGAGKSSWWINPD-AKPGMNP 101

DB 299 NSIRHNLHSRPMRIQNEGAGKSSWWINPD-AKPGMNP 340

RESULT 7

US-08-857-076-102

Sequence 102, Application US/08857076C

Patent No. 6225120

GENERAL INFORMATION:

APPLICANT: Ruvkun, Gary

APPLICANT: Kimura, Koutarou

APPLICANT: Patterson, Garth

APPLICANT: Ogg, Scott

APPLICANT: Paradis, Suzanne

APPLICANT: Tissenbaum, Heidi

APPLICANT: Morris, Jason

APPLICANT: Kowek, Allison

TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR

TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS

FILE REFERENCE: 00786/351001

CURRENT APPLICATION NUMBER: US/08/857,076C

CURRENT FILING DATE: 1997-05-15

NUMBER OF SEQ ID NOS: 114

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 102

LENGTH: 501

TYPE: PRT

ORGANISM: Homo sapiens

US-08-857-076-102

Query Match 73.2%; Score 406.5; DB 3; Length 501;

Best Local Similarity 70.6%; Pred. No. 2.7e-42;

Matches 72; Conservative 15; Mismatches 14; Indels 1; Gaps 1;

QY 1 KKTTRNAGNMSYAEILTTAIMASPEKRLTLAQYVEMVQVYFDRKGDSSAGWK 60

DB 239 RKGGRRNAGNQSYAEILTTAIMASPEKRLTLAQYVEMVQVYFDRKGDSSAGWK 298

QY 61 NSIRHNLHSRPMRIQNEGAGKSSWWINPD-AKPGMNP 101

DB 299 NSIRHNLHSRPMRIQNEGAGKSSWWINPD-AKPGMNP 340

APPLICANT: Stone, Edwin M.
APPLICANT: Nishimura, Darryl
APPLICANT: Patil, Shiva
TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,352
FILING DATE: 22-MAY-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-029.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 99 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-083-352-21
Query Match 70.4%; Score 390.5; DB 3; Length 99;
Best Local Similarity 74.2%; Pred. No. 2.9e-41;
Matches 69; Conservative 15; Mismatches 8; Indels 1; Gaps 1;
Qy 10 WGNMSYAEILITTAIWASPEKRLTLAQVYVWVQVYFRDKGDSNSAGWKNIRHNL 69
Db 1 WGNLSYADLIITKALESSAEKRLTLQIYEWVWVKSVPYFKDKGDSNSAGWKNIRHNL 60
Qy 70 HSRPMRIQNEGAGKSSWWVINPD-AXPGMNP 101
Db 61 HSKIRVQNEGTVGKSSWWMLNPEGGKCKSP 93
RESULT 10
US-08-857-076-46
; Sequence 46, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 509

TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-08-857-076-46
Query Match 68.2%; Score 378.5; DB 3; Length 509;
Best Local Similarity 62.0%; Pred. No. 8.6e-39;
Matches 67; Conservative 15; Mismatches 21; Indels 5; Gaps 1;
Qy 1 KKT-----TRNAGNMSYAEILITTAIWASPEKRLTLAQVYVWVQVYFRDKGDSNS 55
Db 129 KKPTDQLAQKPNPGEESYSDIIAKALESAPDGRKLKLINEIYQWFSNDIPYFGERSPEE 188
Qy 56 SAGWKNIRHNLHLSRPMRIQNEGAGKSSWWVINPDAPKGMNPRTR 103
Db 189 AAGWKNIRHNLHLSRPMRIQNEGAGKSSWWVINPDAPKGMNPRTR 236
RESULT 11
US-08-857-076-101
; Sequence 101, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101
; LENGTH: 635
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-08-857-076-101
Query Match 69.2%; Score 378.5; DB 3; Length 635;
Best Local Similarity 62.0%; Pred. No. 1.2e-38;
Matches 67; Conservative 15; Mismatches 21; Indels 5; Gaps 1;
Qy 1 KKT-----TRNAGNMSYAEILITTAIWASPEKRLTLAQVYVWVQVYFRDKGDSNS 55
Db 255 KKPTDQLAQKPNPGEESYSDIIAKALESAPDGRKLKLINEIYQWFSNDIPYFGERSPEE 314
Qy 56 SAGWKNIRHNLHLSRPMRIQNEGAGKSSWWVINPDAPKGMNPRTR 103
Db 315 AAGWKNIRHNLHLSRPMRIQNEGAGKSSWWVINPDAPKGMNPRTR 362
RESULT 12
US-08-857-076-58
; Sequence 58, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C

```
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-08-857-076-58

Query Match
  Best Local Similarity 68.1%; Score 378; DB 3; Length 98;
  Matches 64; Conservative 14; Mismatches 18; Indels 0; Gaps 0;

QY 8 NAWGNMSYAEIITTAIMASPEKRLTLAQVYEMVQVNYFDRDKGDSNSAGWKNSIRHNL 67
Db 3 NPWGESYSIIAKALESADGRLKNEIYQWFSNDIPIYFGERSPEEAAAGWKNSIRHNL 62

QY 68 SLHSRMRIONEGAGKSSWVWVNPDAKPGNPRTR 103
Db 63 SLHSRMRIONEGAGKSSWVWVNPDAKPGNPRTR 98

RESULT 13
US-09-083-351-16
; Sequence 16, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; TITLE OF INVENTION: TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-083-351-16

Query Match
  Best Local Similarity 34.0%; Score 188.5; DB 3; Length 106;
  Matches 36; Conservative 20; Mismatches 19; Indels 7; Gaps 2;

QY 14 SYAEIITTAIMASPEKRLTLAQVYEMVQVNYFDRDKGDSNSAGWKNSIRHNLHSRFF 73
Db 10 SYALIMMAIRQSPEKRLTLNGIYEFIMKNFPYRE-----NKQGWQNSIRHNLHSRFF 64

QY 74 MRION--EGAGKSSWVWVNPDA 93
Db 65 VKVPRHYDDPGKGNWMLDPSS 86

RESULT 15
US-09-083-351-17
; Sequence 17, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
```

us-09-844-353a-54.rai

```

; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Barryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; TITLE OF INVENTION: TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLRY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-09-083-351-17
;
Query Match 33.4%; Score 185.5; DB 3; Length 106;
Best local Similarity 42.7%; Pred.No.1.3e-15;
Matches 35; Conservative 21; Mismatches 19; Indels 7; Gaps 2;

Qy 14 SYAEILTTATWASPEKILTLTAQYVWVQVNPYPRDKGDSNSAGWKNISIRHNLSLHSRF 73
Db 10 SYNALINWAMRQSEKILTLNGIIEYFIKNEFPYIRE-----NKQGWQNSIRHNLSLNKCF 64
Qy 74 MRIQN--EGAGKSGWVNPDA 93
Db 65 VKVPRHYDDPGKGNVWMLDPSS 86

```

Search completed: July 23, 2004, 10:26:23
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2004, 10:25:31 ; Search time 47 Seconds
(without alignments)
686.274 Million cell updates/sec

Title: US-09-844-353A-54

Perfect score: 555

Sequence: 1 KKTTRRNAGNMYSYAEIIT.....SSWWINPDAPGMNPRTR 103

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	555	100.0	103	9	US-09-205-658-54
2	555	100.0	103	9	US-09-844-353A-54
3	555	100.0	103	10	US-09-963-693-54
4	555	100.0	106	9	US-09-205-658-37
5	555	100.0	106	9	US-09-844-353A-37
6	555	100.0	106	10	US-09-963-693-37
7	555	100.0	109	9	US-09-205-658-56
8	555	100.0	109	9	US-09-844-353A-56
9	555	100.0	109	10	US-09-963-693-56
10	555	100.0	510	9	US-09-205-658-45
11	555	100.0	510	9	US-09-844-353A-45
12	555	100.0	510	10	US-09-963-693-45
13	440.5	79.4	673	16	US-10-701-490-6
14	427.5	77.0	655	9	US-09-205-658-57
15	427.5	77.0	655	9	US-09-844-353A-57

16	427.5	77.0	655	10	US-09-963-693-57	Sequence 57, Appl
17	427.5	77.0	655	15	US-10-341-434-53	Sequence 53, Appl
18	427.5	77.0	655	16	US-10-701-490-3	Sequence 3, Appl
19	410.5	74.0	358	14	US-10-261-517-23	Sequence 23, Appl
20	406.5	73.2	501	9	US-09-844-353A-102	Sequence 102, Appl
21	406.5	73.2	505	14	US-10-186-839-2	Sequence 2, Appl
22	378.5	68.2	509	9	US-09-205-658-46	Sequence 46, Appl
23	378.5	68.2	509	9	US-09-844-353A-46	Sequence 46, Appl
24	378.5	68.2	509	10	US-09-963-693-46	Sequence 46, Appl
25	378.5	68.2	635	9	US-09-844-353A-101	Sequence 101, Appl
26	378	68.1	98	9	US-09-205-658-58	Sequence 58, Appl
27	378	68.1	98	9	US-09-844-353A-58	Sequence 58, Appl
28	378	68.1	98	10	US-09-963-693-58	Sequence 1112, Appl
29	189.5	34.1	190	9	US-09-764-864-1112	Sequence 1530, Appl
30	189.5	34.1	190	9	US-09-764-864-1530	Sequence 2327, A
31	182.5	32.9	237	14	US-10-029-386-32327	Sequence 136, Appl
32	182.5	32.9	465	14	US-10-205-823-136	Sequence 2, Appl
33	182.5	32.9	553	10	US-09-292-862-2	Sequence 2, Appl
34	182	32.8	748	13	US-10-151-587-2	Sequence 7, Appl
35	182	32.8	748	16	US-10-650-609-2	Sequence 2, Appl
36	181.5	32.7	494	9	US-09-963-285-7	Sequence 7, Appl
37	181.5	32.7	501	9	US-09-963-285-2	Sequence 2, Appl
38	181.5	32.7	501	9	US-09-963-285-10	Sequence 10, Appl
39	179	32.3	84	12	US-10-150-813-78	Sequence 78, Appl
40	179	32.3	84	12	US-10-139-854-78	Sequence 78, Appl
41	179	32.3	84	14	US-10-131-409-78	Sequence 78, Appl
42	179	32.3	84	15	US-10-150-811-78	Sequence 78, Appl
43	179	32.3	335	9	US-09-764-864-1178	Sequence 1178, Appl
44	179	32.3	464	14	US-10-148-662-10	Sequence 10, Appl
45	179	32.3	489	14	US-10-148-662-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1

US-09-205-658-54
; Sequence 54, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-205-658-54

Query Match 100.0%; Score 555; DB 9; Length 103;
Best Local Similarity 100.0%; Pred. No. 3.4e-58;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKTTRRNAGNMYSYAEIITAINASPEKRLTIAQVYEWVQNVYFRDKGDSNAGWK 60
Db 1 KKTTRRNAGNMYSYAEIITAINASPEKRLTIAQVYEWVQNVYFRDKGDSNAGWK 60

Qy 61 NSIRHNLSLSRFRMIRIQNEGAGKSSWWINPDAPGMNPRTR 103
Db 61 NSIRHNLSLSRFRMIRIQNEGAGKSSWWINPDAPGMNPRTR 103


```
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-844-353A-37

Query Match      100.0%; Score 555; DB 9; Length 106;
Best Local Similarity 100.0%; Pred. No. 3.5e-58;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKTTRRNAGNMSYAEIITTAIMASPEKRLTLAQYEWVQVVPYFRDKGDSNSAGWK 60
Db 1 KKTTRRNAGNMSYAEIITTAIMASPEKRLTLAQYEWVQVVPYFRDKGDSNSAGWK 60

QY 61 NSIRHNLHSRFRMRIQNEGAGKSSWWVINPDAPGMMNPRTR 103
Db 61 NSIRHNLHSRFRMRIQNEGAGKSSWWVINPDAPGMMNPRTR 103

RESULT 6
US-09-963-693-37
; Sequence 37, Application US/09963693
; Publication No. US20030181364A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/963,693
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/205,658
; PRIOR FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 08/857,076
; PRIOR FILING DATE: 1997-05-15
; PRIOR APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-963-693-37

Query Match      100.0%; Score 555; DB 10; Length 106;
Best Local Similarity 100.0%; Pred. No. 3.5e-58;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKTTRRNAGNMSYAEIITTAIMASPEKRLTLAQYEWVQVVPYFRDKGDSNSAGWK 60
Db 1 KKTTRRNAGNMSYAEIITTAIMASPEKRLTLAQYEWVQVVPYFRDKGDSNSAGWK 60

QY 61 NSIRHNLHSRFRMRIQNEGAGKSSWWVINPDAPGMMNPRTR 103
Db 61 NSIRHNLHSRFRMRIQNEGAGKSSWWVINPDAPGMMNPRTR 103

RESULT 7
US-09-205-658-56
; Sequence 56, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
```

```
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-205-658-56

Query Match      100.0%; Score 555; DB 9; Length 109;
Best Local Similarity 100.0%; Pred. No. 3.6e-58;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKTTRRNAGNMSYAEIITTAIMASPEKRLTLAQYEWVQVVPYFRDKGDSNSAGWK 60
Db 7 KKTTRRNAGNMSYAEIITTAIMASPEKRLTLAQYEWVQVVPYFRDKGDSNSAGWK 66

QY 61 NSIRHNLHSRFRMRIQNEGAGKSSWWVINPDAPGMMNPRTR 103
Db 67 NSIRHNLHSRFRMRIQNEGAGKSSWWVINPDAPGMMNPRTR 109

RESULT 8
US-09-844-353A-56
; Sequence 56, Application US/09844353A
; Patent No. US20020037585A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowsek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351005
; CURRENT APPLICATION NUMBER: US/09/844,353A
; CURRENT FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 08/857,076
; PRIOR FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-844-353A-56

Query Match      100.0%; Score 555; DB 9; Length 109;
Best Local Similarity 100.0%; Pred. No. 3.6e-58;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKTTRRNAGNMSYAEIITTAIMASPEKRLTLAQYEWVQVVPYFRDKGDSNSAGWK 60
Db 7 KKTTRRNAGNMSYAEIITTAIMASPEKRLTLAQYEWVQVVPYFRDKGDSNSAGWK 66

QY 61 NSIRHNLHSRFRMRIQNEGAGKSSWWVINPDAPGMMNPRTR 103
Db 67 NSIRHNLHSRFRMRIQNEGAGKSSWWVINPDAPGMMNPRTR 109

RESULT 9
US-09-963-693-56
; Sequence 56, Application US/09963693
; Publication No. US20030181364A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
```

```

; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/963,693
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/205,658
; PRIOR FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 08/857,076
; PRIOR FILING DATE: 1997-05-15
; PRIOR APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 56
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-963-693-56

Query Match      100.0%; Score 555; DB 10; Length 109;
Best Local Similarity 100.0%; Pred. No. 3.6e-58;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 KKTTRNANGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFPRDKGDSNSSAGWK 60
Db      7 KKTTRNANGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFPRDKGDSNSSAGWK 66

Qy      61 NSIRHNLSLHSRFRMRIONEAGKSSWWVINPDAPKGMNPRTR 103
Db      67 NSIRHNLSLHSRFRMRIONEAGKSSWWVINPDAPKGMNPRTR 109

RESULT 10
US-09-205-658-45
; Sequence 45, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-205-658-45

Query Match      100.0%; Score 555; DB 9; Length 510;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 KKTTRNANGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFPRDKGDSNSSAGWK 60
Db      135 KKTTRNANGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFPRDKGDSNSSAGWK 194

Qy      61 NSIRHNLSLHSRFRMRIONEAGKSSWWVINPDAPKGMNPRTR 103
Db      195 NSIRHNLSLHSRFRMRIONEAGKSSWWVINPDAPKGMNPRTR 237

RESULT 11
US-09-205-658-45
; Sequence 45, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-205-658-45

Query Match      100.0%; Score 555; DB 9; Length 510;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 KKTTRNANGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFPRDKGDSNSSAGWK 60
Db      135 KKTTRNANGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFPRDKGDSNSSAGWK 194

Qy      61 NSIRHNLSLHSRFRMRIONEAGKSSWWVINPDAPKGMNPRTR 103
Db      195 NSIRHNLSLHSRFRMRIONEAGKSSWWVINPDAPKGMNPRTR 237

RESULT 12
US-09-963-693-45
; Sequence 45, Application US/09963693
; Publication No. US20030181364A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/963,693
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/205,658
; PRIOR FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 08/857,076
; PRIOR FILING DATE: 1997-05-15
; PRIOR APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-963-693-45

Query Match      100.0%; Score 555; DB 10; Length 510;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 KKTTRNANGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFPRDKGDSNSSAGWK 60
Db      135 KKTTRNANGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFPRDKGDSNSSAGWK 194

Qy      61 NSIRHNLSLHSRFRMRIONEAGKSSWWVINPDAPKGMNPRTR 103
Db      195 NSIRHNLSLHSRFRMRIONEAGKSSWWVINPDAPKGMNPRTR 237

US-09-844-353A-45
; Sequence 45, Application US/09844353A
; Patent No. US20020037585A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowsek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351005
; CURRENT APPLICATION NUMBER: US/09/844,353A
; CURRENT FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 08/857,076
; PRIOR FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-844-353A-45

Query Match      100.0%; Score 555; DB 9; Length 510;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 KKTTRNANGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFPRDKGDSNSSAGWK 60
Db      135 KKTTRNANGNMSYAEIITTAIMASPEKRLTLAQVYEMVQNVYFPRDKGDSNSSAGWK 194

Qy      61 NSIRHNLSLHSRFRMRIONEAGKSSWWVINPDAPKGMNPRTR 103
Db      195 NSIRHNLSLHSRFRMRIONEAGKSSWWVINPDAPKGMNPRTR 237
```

Db 135 KKTTRNANGNSYAEIITTAIMASPEKRLTLAQVYEWVQNVYFRDKGDSNSSAGWK 194
QY 61 NSIRHNLHLSRPMRIQNEGAGKSSWWVINPDAPGMNPRTR 103
Db 195 NSIRHNLHLSRPMRIQNEGAGKSSWWVINPDAPGMNPRTR 237

RESULT 13
US-10-701-490-6
; Sequence 6, Application US/10701490
; Publication No. US20040106141A1
; GENERAL INFORMATION:
; APPLICANT: PAUL S. MISCHER
; APPLICANT: CHARLES L. SAWYERS
; APPLICANT: BRADLEY L. SMITH
; APPLICANT: KATHERINE CROSBY
; TITLE OF INVENTION: METHODS AND MATERIALS FOR EXAMINING
; FILE OF INVENTION: PATHWAYS ASSOCIATED WITH GLOBLASTOMA PROGRESSION
; FILE REFERENCE: G&C 30435.148USU1
; CURRENT APPLICATION NUMBER: US/10/701,490
; PRIOR FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: 60/423,777
; PRIOR FILING DATE: 2002-11-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 673
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-701-490-6

Query Match 79.4%; Score 440.5; DB 16; Length 673;
Best Local Similarity 77.5%; Pred. No. 1.5e-43;
Matches 79; Conservative 13; Mismatches 9; Indels 1; Gaps 1;
QY 1 KKTTRNANGNSYAEIITTAIMASPEKRLTLAQVYEWVQNVYFRDKGDSNSSAGWK 60
Db 148 KCSSRRNANGNSYADLITKAIESSAEKRLTLQSIYEWVWVCVPYFKDKGDSNSSAGWK 207
QY 61 NSIRHNLHLSRPMRIQNEGAGKSSWWVINPD-APGMNPRR 101
Db 208 NSIRHNLHLSRPMRIQNEGAGKSSWWVINPDGKGSKAPRR 249

RESULT 14
US-09-205-658-57
; Sequence 57, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; PRIOR FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; PRIOR FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-205-658-57

Query Match 77.0%; Score 427.5; DB 9; Length 655;
Best Local Similarity 73.5%; Pred. No. 5.1e-42;

Matches 75; Conservative 18; Mismatches 8; Indels 1; Gaps 1;
QY 1 KKTTRNANGNSYAEIITTAIMASPEKRLTLAQVYEWVQNVYFRDKGDSNSSAGWK 60
Db 151 KSSSRNANGNSYADLITKAIESSAEKRLTLQSIYEWVWVCVPYFKDKGDSNSSAGWK 210
QY 61 NSIRHNLHLSRPMRIQNEGAGKSSWWVINPD-APGMNPRR 101
Db 211 NSIRHNLHLSRPMRIQNEGAGKSSWWVINPDGKGSKAPRR 252

RESULT 15
US-09-844-353A-57
; Sequence 57, Application US/09844353A
; Patent No. US20020037585A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351005
; CURRENT APPLICATION NUMBER: US/09/844,353A
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 08/857,076
; PRIOR FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-844-353A-57

Query Match 77.0%; Score 427.5; DB 9; Length 655;
Best Local Similarity 73.5%; Pred. No. 5.1e-42;
Matches 75; Conservative 18; Mismatches 8; Indels 1; Gaps 1;
QY 1 KKTTRNANGNSYAEIITTAIMASPEKRLTLAQVYEWVQNVYFRDKGDSNSSAGWK 60
Db 151 KSSSRNANGNSYADLITKAIESSAEKRLTLQSIYEWVWVCVPYFKDKGDSNSSAGWK 210
QY 61 NSIRHNLHLSRPMRIQNEGAGKSSWWVINPD-APGMNPRR 101
Db 211 NSIRHNLHLSRPMRIQNEGAGKSSWWVINPDGKGSKAPRR 252

Search completed: July 23, 2004, 10:31:01
Job time : 47 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2004, 10:22:15 ; Search time 16 Seconds
(without alignments)
619.233 Million cell updates/sec

Title: US-09-844-353A-54
Perfect score: 555
Sequence: 1 KKTTRRNAWGNMSYAEIIT.....SSWWINPDAPKGMNPRTR 103

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	549	98.9	510	2 T42255	HNF-3/forkhead pro
2	427.5	77.0	655	2 S40521	FKHR protein - hum
3	372.5	67.1	530	2 T42234	fork head-related
4	372.5	67.1	622	2 T37204	hypothetical prote
5	322	58.0	235	2 T37203	hypothetical prote
6	209.5	37.7	230	2 T19437	hypothetical prote
7	195.5	35.2	270	2 T16880	hypothetical prote
8	191	34.4	310	2 T33497	hypothetical prote
9	190	34.2	101	2 A47450	HNF-3/fork-head ho
10	188.5	34.0	109	2 B54743	transcription fact
11	188.5	34.0	387	2 A47446	HNF-3/fork head fa
12	188.5	34.0	451	2 A55909	transforming prote
13	188.5	34.0	476	2 A54743	transcription fact
14	188.5	34.0	480	2 JH0672	brain factor 1 pro
15	186	33.5	445	2 S23055	Slp2 protein - fru
16	186	33.5	445	2 S30556	Slp2 protein - fru
17	185.5	33.4	469	2 I37451	HBF-G2 (HPK-2) pro
18	185	33.3	134	2 B46178	probable transcrip
19	184.5	33.2	532	2 S22262	transcription fact
20	184	33.2	101	2 I60919	HNF-3/fork-head ho
21	183.5	33.1	586	2 J65000	hnf-3/forkhead tra
22	183.5	33.1	663	2 T40493	hnf-3/forkhead tra
23	182.5	32.9	101	2 I60922	brain factor-2 - r
24	182.5	32.9	106	2 S51627	forkhead protein F
25	182.5	32.9	106	2 S51626	FKRAC-3 protein -
26	182.5	32.9	117	2 I49674	transcription fact
27	182.5	32.9	440	2 S71795	transcription fact
28	182.5	32.9	465	2 G02738	FKRAC-4 - human
29	181.5	32.7	461	2 S34472	MFH-1 protein - mo

30 179.5 32.3 320 2 A42826 T-cell leukemia vi
31 179.5 32.3 322 2 S23053 sloppy paired prot
32 179 32.3 84 2 A49395 glutamine-rich fac
33 179 32.3 323 2 T15311 hypothetical prote
34 178.5 32.2 543 2 A41285 interleukin enhanc
35 177.5 32.0 101 2 I60917 HNF-3/fork-head ho
36 177.5 32.0 112 2 C54743 transcription fact
37 176.5 31.8 106 2 S51624 FREAC-1 protein - Af
38 176.5 31.8 427 2 I51580 XFKH2 protein - Af
39 176.5 31.8 466 2 A36674 transcription fact
40 176 31.7 400 2 A42377 fork head domain p
41 175.5 31.6 101 2 I60918 brain factor-3 - r
42 175.5 31.6 111 2 E56556 fork head homolog
43 175.5 31.6 468 1 A54258 transcription fact
44 175.5 31.6 473 2 S70357 forkhead transcrip
45 175.5 31.6 508 2 S59870 fork head domain p

ALIGNMENTS

RESULT 1

T42255
HNF-3/forkhead protein homolog daf-16 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 24-Aug-2001
C:Accession: T42255
R:Lin, K.; Dorman, J.B.; Rodan, A.; Kenyon, C.
Science 278, 1319-1322, 1997
A:Title: Daf-16, An HNF-3/forkhead family member that can function to double the life span of C. elegans
A:Reference number: Z22130; MUID:98028757; PMID:9360933
A:Accession: T42255
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-510 <LIN>
A:Cross-references: EMBL:AF032112; NID:g2623942; PIDN:AAC47803.1; PID:g2623943
C:Genetics:
A:Gene: daf-16
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

Query Match 98.9%; Score 549; DB 2; Length 510;
Best Local Similarity 99.0%; Pred. No. 3.4e-49;
Matches 102; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KKTTRRNAWGNMSYAEIITTAIWASPEKRLTAQVYEWVQNVYFRDKGDSNSAGWK 60
Db 135 KKTTRRNAWGNMSYAEIITTAIWASPEKRLTAQVYEWVQNVYFRDKGDSNSAGWK 194
QY 61 NSIRHNLHGRFMRIQNEGAGKGSWWVINPDAPKGMNPRTR 103
Db 195 NSIRHNLHGRFMRIQNEGAGKGSWWVINPDAPKGMNPRTR 237

RESULT 2

S40521
FKHR protein - human
C:Species: Homo sapiens (man)
C:Date: 07-Oct-1994 #sequence_revision 03-Nov-1995 #text_change 07-May-1999
C:Accession: S40521
R:Galili, N.; Davis, R.J.; Fredericks, W.J.; Mukhopadhyay, S.; Rauscher III, F.J.; Emanuel
Nature Genet. 5, 230-235, 1993
A:Title: Fusion of a fork head domain gene to PAX3 in the solid tumour alveolar rhabdomyosarcoma
A:Reference number: S40521; MUID:94100975; PMID:8275086
A:Accession: S40521
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-655 <GAL>
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F:160-248/Domain: fork head DNA-binding domain homology <FHD>

Query Match 77.0%; Score 427.5; DB 2; Length 655;
Best Local Similarity 73.5%; Pred. No. 1.8e-36;
Matches 75; Conservative 18; Mismatches 8; Indels 1; Gaps 1;

Db 210 AAGWKNISIRHNLHLSRFRMR1QNEGAGKSSWWVNPDAKPGNPRRTR 257

RESULT 5

T37203

hypothetical protein R13H8.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T37203

R:Jones, K.; Hinds, K.; Sutterer, C.; Cofman, M.

A:Description: The sequence of C. elegans cosmid R13H8.

A:Reference number: Z21633

A:Accession: T37203

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-235 <ON>

A:Cross-references: EMBL:AF039717; PIDN:AAB96742.1

A:Experimental source: strain Bristol N2

C:Genetics:

A:Map position: 1

A:Introns: 82/2; 121/2; 168/1

A:Note: R13H8.2

Query Match 58.0%; Score 322; DB 2; Length 235;

Best Local Similarity 96.8%; Pred. No. 5.1e-26;

Matches 61; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQYVWVQVYPRDKGDSNSAGWK 60

Db 133 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQYVWVQVYPRDKGDSNSAGWK 192

QY 61 NSI 63

Db 193 VSL 195

RESULT 6

T19437

hypothetical protein C25A1.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T19437

R:Mortimore, B.

A:Submitted to the EMBL Data Library, October 1996

A:Reference number: Z19124

A:Accession: T19437

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-230 <WIL>

A:Cross-references: EMBL:Z81038; PIDN:CAR02761.1; GSPDB:GN00019; CESP:C25A1.2

A:Experimental source: clone C25A1

C:Genetics:

A:Gene: CESP:C25A1.2

A:Map position: 1

A:Introns: 35/2; 103/3; 146/3

Query Match 37.7%; Score 209.5; DB 2; Length 230;

Best Local Similarity 47.4%; Pred. No. 2.4e-14;

Matches 37; Conservative 16; Mismatches 20; Indels 5; Gaps 1;

QY 14 SYAEIITTAIMASPEKRLTLAQYVWVQVYPRDKGDSNSAGWKNSIRHNLHLSRFR 73

Db 82 SYGLIAAILSSPQKMWLAEVYEWIMNEYVYFRSRG-----AGWRNSIRHNLSDNCF 136

QY 74 MR1QNEGAGKSSWWVNP 91

Db 137 VKAGRAANGKGYWAVHP 154

RESULT 7

T16980

hypothetical protein T14G12.4 - Caenorhabditis elegans

QY 1 KKTTRNAGNMSYAEIITTAIMASPEKRLTLAQYVWVQVYPRDKGDSNSAGWK 60

Db 151 KSSSRNAGNLSYADLTITKAIESAEKRLTLAQYVWVQVYPRDKGDSNSAGWK 210

QY 61 NSIRHNLHLSRFRMR1QNEGAGKSSWWVNPDAKPGNPRR 101

Db 211 NSIRHNLHLSRFRMR1QNEGAGKSSWWVNPDAKPGNPRR 252

RESULT 3

T42234

fork head-related transcription factor homolog - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000

C:Accession: T42234

R:Ogg, S.; Paradis, S.; Gottlieb, S.; Patterson, G.I.; Lee, L.; Tissenbaum, H.A.; Ruvkun

A:Title: The Fork head transcription factor DAF-16 transduces insulin-like metabolic and

A:Reference number: Z22108; MUID:98013175; PMID:9353126

A:Accession: T42234

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-530 <OGG>

A:Cross-references: EMBL:AF020344; NID:G2618980; PIDN:AAB84392.1; PID:G2618981

C:Genetics:

A:Gene: daf-16

C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

Query Match 67.1%; Score 372.5; DB 2; Length 530;

Best Local Similarity 61.1%; Pred. No. 7.4e-31;

Matches 66; Conservative 15; Mismatches 22; Indels 5; Gaps 1;

QY 1 KKT-----TNRNAGNMSYAEIITTAIMASPEKRLTLAQYVWVQVYPRDKGDSNS 55

Db 150 KKPTDQLAQKPNPWEESYSDIAKALAPDGLKNEIYQWFSNIPYFGERSPEE 209

QY 56 SAGWKNISIRHNLHLSRFRMR1QNEGAGKSSWWVNPDAKPGNPRRTR 103

Db 210 AAGWKNISIRHNLHLSRFRMR1QNEGAGKSSWWVNPDAKPGNPRRTR 257

RESULT 4

T37204

hypothetical protein R13H8.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Jun-2000

C:Accession: T37204

R:Jones, K.; Hinds, K.; Sutterer, C.; Cofman, M.

A:Submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid R13H8.

A:Reference number: Z21633

A:Accession: T37204

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-822 <WIL>

A:Cross-references: EMBL:AF039717; PIDN:AAB96741.1

C:Genetics:

A:Map position: 1

A:Introns: 214/3; 268/3; 329/2; 419/3; 457/2; 547/2; 563/1

A:Note: R13H8.1

C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

Query Match 67.1%; Score 372.5; DB 2; Length 622;

Best Local Similarity 61.1%; Pred. No. 8.9e-31;

Matches 66; Conservative 15; Mismatches 22; Indels 5; Gaps 1;

QY 1 KKT-----TNRNAGNMSYAEIITTAIMASPEKRLTLAQYVWVQVYPRDKGDSNS 55

Db 150 KKPTDQLAQKPNPWEESYSDIAKALAPDGLKNEIYQWFSNIPYFGERSPEE 209

QY 56 SAGWKNISIRHNLHLSRFRMR1QNEGAGKSSWWVNPDAKPGNPRRTR 103

A;Title: Identification of nine tissue-specific transcription factors of the hepatocyte

A;Reference number: A47450; MUID:93248207; PMID:7683413

A;Accession: A47450

A>Status: preliminary; translated from GB/EMBL/DDBB

A:Molecule type: mRNA

A;Residues: 1-101 <RES>

A;Cross-references: GDB:L13203; NID:g306843; PID:g306844

C;Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

C;Keywords: DNA binding; transcription factor

F;3-94/Domain: fork head DNA-binding domain homology <FHD>

Query Match 34.2%; Score 190; DB 2; Length 101;
Best Local Similarity 40.0%; Pred. No. 9.6e-13;
Matches 38; Conservative 22; Mismatches 25; Indels 10; Gaps 3;

QY 14 SYAELITTAIMASPEKRLTLAQVYEWVNVPYFRDKGDSNSAGWKNSIRHNLSLHSRF 73
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Db 7 SYSALIAAIHAGPKRUTLSQTYYADNFFFY-----NKSKAGWQNSIRHNLSLNDKF 61

QY 74 MRI--QNEGAGKSSWWVINPDAK---PGMNPRTTR 103
:::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Db 62 KKVPREDDEDPCKGNVTLDPCNEKMFENRRKR 96

RESULT 10

B54743

transcription factor HFK2 - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 12-Sep-1997

C;Accession: B54743

R;Murphy, D.B.; Wieser, S.; Burfeind, P.; Schmundt, D.; Mattei, M.G.; Schulz-Schaeffer, W.

G;Genomics 21, 551-557, 1994

A;Title: Human brain factor 1, a new member of the fork head gene family.

A;Reference number: A54743; MUID:95048332; PMID:7959731

A;Accession: B54743

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A;Residues: 1-109 <MUR>

A;Cross-references: GDB:X74143

C;Genetics:

A;Gene: GDB:FHL2; HBF-2; HFK2

A;Cross-references: GDB:I375747; OMIM:600779

A;Map position: 14q11-14q13

C;Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

F;8-99/Domain: fork head DNA-binding domain homology <FHD>

Query Match 34.0%; Score 188.5; DB 2; Length 109;
Best Local Similarity 43.9%; Pred. No. 1.5e-12;
Matches 36; Conservative 20; Mismatches 19; Indels 7; Gaps 2;

QY 14 SYAELITTAIMASPEKRLTLAQVYEWVNVPYFRDKGDSNSAGWKNSIRHNLSLHSRF 73
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Db 12 SYNPLIWAIRQSPKRLTNGIEFMKNFPFYRE----NKQGWNQSIRHNLSLNKKCF 66

QY 74 MRION--EGAGKSSWWVINPDA 93
:::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Db 67 VKVPRHYDDPGKGNVWLDPSS 88

RESULT 11

A47446

HNF-3/fork head family transcription factor Qin - avian sarcoma virus 31

C;Species: avian sarcoma virus 31, ASV31

C;Date: 03-Mar-1994 #sequence_revision 25-Apr-1997 #text_change 19-May-2000

C;Accession: A47446

R;Li, J.; Vogt, P.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 4490-4494, 1993

A;Title: The retroviral oncogene qin belongs to the transcription factor family that inc

A;Reference number: A47446; MUID:93281605; PMID:8099441

A;Accession: A47446

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A;Residues: 1-387 <LIJ>

A;Cross-references: GB:L10719; NID:g303962
A;Note: sequence extracted from NCBI backbone (NCBIP:133114)
A;Note: this translation is not annotated in GenBank entry ACSTRANS, release 113.0
C;Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
C;Keywords: transcription factor
F:151-242/Domain: fork head DNA-binding domain homology <FHD>

Query Match 34.0%; Score 188.5; DB 2; Length 387;
Best Local Similarity 43.9%; Pred. No. 6.5e-12;
Matches 36; Conservative 20; Mismatches 19; Indels 7; Gaps 2;

QY 14 SYAELITTAIMASPEKRLTLAQVYEMVQNVYPRDKGDSNSAGWKNISIRHNLHLHSRF 73
DB 155 SYNALIMMAIROSPKRLTLNGIYEFIMKNFPYIRE-----NKQGWQNSIRHNLHLKCF 209

QY 74 MRION--EGAGKSSWWVINPDA 93
DB 210 VKVPRHYDDPGKGYWMLDPSS 231

RESULT 12
A55909
transforming protein c-qin - chicken
N;Alternate names: transcription factor CBF-1
C;Species: Gallus gallus (chicken)
C;Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 24-Sep-1999
C;Accession: A55909; S71794
R;Chang, H.W.; Li, J.; Kretzschmar, D.; Vogt, P.K.
Proc. Natl. Acad. Sci. U.S.A. 92, 447-451, 1995
A;Title: Avian cellular homolog of the c-in oncogene.
A;Reference number: A55909; MUID:95132616; PMID:7831308
A;Accession: A55909
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-451 <CHA>
A;Cross-references: GB:L36814; NID:g642602; PIDN:AAA66954.1; PID:g642603
R;Yuasa, J.; Hirano, S.; Yamagata, M.; Noda, M.
Nature 382, 632-635, 1996
A;Title: Visual projection map specified by topographic expression of transcription factor
A;Reference number: S71794; MUID:96338226; PMID:8757134
A;Accession: S71794
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-451 <YUA>
A;Cross-references: EMBL:U47275; NID:gl546781; PIDN:AB08466.1; PID:gl546782
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1996
C;Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
C;Keywords: DNA binding; proto-oncogene
F:143-234/Domain: fork head DNA-binding domain homology <FHD>

Query Match 34.0%; Score 188.5; DB 2; Length 451;
Best Local Similarity 43.9%; Pred. No. 7.8e-12;
Matches 36; Conservative 20; Mismatches 19; Indels 7; Gaps 2;

QY 14 SYAELITTAIMASPEKRLTLAQVYEMVQNVYPRDKGDSNSAGWKNISIRHNLHLHSRF 73
DB 147 SYNALIMMAIROSPKRLTLNGIYEFIMKNFPYIRE-----NKQGWQNSIRHNLHLKCF 201

QY 74 MRION--EGAGKSSWWVINPDA 93
DB 202 VKVPRHYDDPGKGYWMLDPSS 223

RESULT 13
A54743
transcription factor HFK1 - human
C;Species: Homo sapiens (man)
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 12-Sep-1997
C;Accession: A54743
R;Murphy, D.B.; Wiese, S.; Burfeind, P.; Schmundt, D.; Mattei, M.G.; Schulz-Schaeffer, W.
Genomics 21, 551-557, 1994
A;Title: Human brain factor 1, a new member of the fork head gene family.
A;Reference number: A54743; MUID:95048332; PMID:7959731

A;Accession: A54743
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-476 <MUR>
A;Cross-references: GB:X74142
C;Genetics:
A;Gene: GDB:FKHL4; HBP-1; HFK1
A;Cross-references: GDB:433550
A;Map position: 14q12-14q12
C;Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F:169-260/Domain: fork head DNA-binding domain homology <FHD>

Query Match 34.0%; Score 188.5; DB 2; Length 476;
Best Local Similarity 43.9%; Pred. No. 8.2e-12;
Matches 36; Conservative 20; Mismatches 19; Indels 7; Gaps 2;

QY 14 SYAELITTAIMASPEKRLTLAQVYEMVQNVYPRDKGDSNSAGWKNISIRHNLHLHSRF 73
DB 173 SYNALIMMAIROSPKRLTLNGIYEFIMKNFPYIRE-----NKQGWQNSIRHNLHLKCF 227

QY 74 MRION--EGAGKSSWWVINPDA 93
DB 228 VKVPRHYDDPGKGYWMLDPSS 249

RESULT 14
JH0672
brain factor 1 protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 24-Sep-1999
C;Accession: JH0672
R;Tao, W.; Lai, E.
Neuron 8, 957-966, 1992
A;Title: Telencephalon-restricted expression of BF-1, a new member of the HNF-3/fork head
A;Reference number: JH0672; MUID:92265309; PMID:1350202
A;Accession: JH0672
A;Molecule type: mRNA
A;Residues: 1-480 <TAO>
A;Cross-references: GB:M87634; NID:g203134; PIDN:AAA40812.1; PID:g203135
A;Experimental source: brain
C;Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
C;Keywords: DNA binding; transcription factor
F:162-271/Region: DNA binding #status predicted
F:172-263/Domain: fork head DNA-binding domain homology <FHD>

Query Match 34.0%; Score 188.5; DB 2; Length 480;
Best Local Similarity 43.9%; Pred. No. 8.3e-12;
Matches 36; Conservative 20; Mismatches 19; Indels 7; Gaps 2;

QY 14 SYAELITTAIMASPEKRLTLAQVYEMVQNVYPRDKGDSNSAGWKNISIRHNLHLHSRF 73
DB 176 SYNALIMMAIROSPKRLTLNGIYEFIMKNFPYIRE-----NKQGWQNSIRHNLHLKCF 230

QY 74 MRION--EGAGKSSWWVINPDA 93
DB 231 VKVPRHYDDPGKGYWMLDPSS 252

RESULT 15
S23055
Slp2 protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 19-Feb-1994 #sequence_revision 01-Sep-1995 #text_change 24-Sep-1999
C;Accession: S23055
R;Grossniklaus, U.; Pearson, R.K.; Gehring, W.J.
Genes Dev. 6, 1030-1051, 1992
A;Title: The Drosophila sloppy paired locus encodes two proteins involved in segmentation
A;Reference number: S23053; MUID:92275347; PMID:1317319
A;Accession: S23055
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-445 <GRO>
A;Cross-references: EMBL:X66098; NID:g8620; PIDN:CAA46892.1; PID:g8621

Search completed: July 23, 2004, 10:25:54
Job time : 17 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2004, 10:21:10 ; Search time 13 Seconds
(without alignments)
412.556 Million cell updates/sec

Title: US-09-844-353A-54

Perfect score: 555

Sequence: 1 KKTTRRNGNMSYAEIIT.....SSWVINPDAGGNPRTR 103

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	440.5	79.4	673	1	FXO3 HUMAN
2	427.5	77.0	655	1	FXO1 HUMAN
3	426.5	76.8	652	1	FXO1 MOUSE
4	413.5	74.5	505	1	FXO4 MOUSE
5	410.5	74.0	505	1	FXO4 HUMAN
6	222.5	40.1	622	1	YK41 HUMAN
7	201.5	36.3	565	1	FXJ2 MOUSE
8	201.5	36.3	574	1	FXJ2 HUMAN
9	190	34.2	351	1	FXI1 HUMAN
10	190	34.2	743	1	YK78 SCHPO
11	188.5	34.0	387	1	QIN AVIS3
12	188.5	34.0	451	1	FXGB CHICK
13	188.5	34.0	477	1	FXGB HUMAN
14	188.5	34.0	480	1	FXGB RAT
15	188.5	34.0	481	1	FXGB MOUSE
16	187.5	33.8	365	1	FD2 DROME
17	186	33.5	445	1	SLP2 DROME
18	185.5	33.4	469	1	FXGA HUMAN
19	184.5	33.2	456	1	FXD1 MOUSE
20	184	33.2	101	1	FXI1 RAT
21	183.5	33.1	663	1	SEPI SCHPO
22	182.5	32.9	101	1	FXGA RAT
23	182.5	32.9	440	1	FXGA CHICK
24	182.5	32.9	465	1	FXD1 HUMAN
25	182.5	32.9	497	1	FXD2 HUMAN
26	182.5	32.9	553	1	FXC1 MOUSE
27	182.5	32.9	553	1	FXC1 HUMAN
28	182	32.8	763	1	FXM1 HUMAN
29	181.5	32.7	494	1	FXC2 MOUSE
30	181.5	32.7	501	1	FXC2 HUMAN
31	180.5	32.5	759	1	FXM1 RAT
32	180.5	32.5	760	1	FXM1 MOUSE
33	180	32.4	490	1	FXN3 HUMAN

34	179.5	32.3	322	1	SLP1 DROME
35	179.5	32.3	341	1	FXN2 HUMAN
36	179.5	32.3	354	1	FXF1 HUMAN
37	179.5	32.3	564	1	HCM1 YEAST
38	179	32.3	677	1	FXP1 HUMAN
39	179	32.3	705	1	FXP1 MOUSE
40	178.5	32.2	655	1	ILF1 HUMAN
41	178	32.1	376	1	FXL2 HUMAN
42	177.5	32.0	101	1	FXD3 RAT
43	177.5	32.0	224	1	FXGC HUMAN
44	177.5	32.0	394	1	FXD3 CHICK
45	177.5	32.0	415	1	HN3B ORYLA

ALIGNMENTS

RESULT 1
FXO3_HUMAN
ID AC FXO3_HUMAN STANDARD; PRT; 673 AA.
AC O43524; O15171; Q9BZ04;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Forkhead box protein O3A (Forkhead in rhabdomyosarcoma-like 1) (AF6q21
DE protein).
GN FOXO3A OR FKHL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98140118; PubMed=9479491;
RA Anderson M.J., Viars C.S., Czekay S., Cavenee W.K., Arden K.C.;
RT "Cloning and characterization of three human forkhead genes that
RT comprise an FKHR-like gene subfamily.",
RL Genomics 47:187-199(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Cobley V.,
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krywinski M.I., Skalski U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.",
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 1-383 FROM N.A.
RX MEDLINE=98008138; PubMed=9345057;
RA Hillion J., Le Coniat M., Jonveaux P., Berger R., Bernard O.A.;
RT "AP6q21, a novel partner of the MLL gene in t(6;11)(q21;q23), defines
RT a forkhead transcriptional factor subfamily.",
RL Blood 90:3714-3719(1997).

Matches 79; Conservative 13; Mismatches 9; Indels 1; Gaps 1;

QY 1 KKTTRNRNAGNMSYAEIITTAIMASPEKRLTLAQVYEMVQVNYFYFDKDGDSNSAGWK 60
 Db 148 KCSRRNAGNLSYADLITRAIESSPDKRLTSLQIYEMVRCVYFDKDGDSNSAGWK 207
 QY 61 NSIRHLSLHSPMRQNEGAGCKSSWWVNPDP-APGGMPPR 101
 Db 208 NSIRHLSLHSPMRQNEGAGCKSSWWVNPDPGKSGKAPRR 249

RESULT 2

FX01 HUMAN
 ID -FX01 HUMAN STANDARD; PRT; 655 AA.
 AC Q12778; O43523;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Forkhead box protein O1A (Forkhead in rhabdomyosarcoma).
 GN FOXO1A OR FKHR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=94100975; PubMed=8275086;
 RA Galili N., Davis R.J., Fredericks W.J., Mukhopadhyay S.,
 RA Rauscher F.J. III, Emanuel B.S., Rovera G., Barr F.G.;
 RA "Fusion of a fork head domain gene to PAX3 in the solid tumour
 RT alveolar rhabdomyosarcoma.";
 RL Nat. Genet. 5:230-235(1993).
 [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98140118; PubMed=9479491;
 RA Anderson M.J., Viars C.S., Czekay S., Cavenee W.K., Arden K.C.;
 RA "Cloning and characterization of three human forkhead genes that
 RT comprise an FKHR-like gene subfamily.";
 RL Genomics 47:187-199(1998).
 [3]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Lymph.
 RC MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences.";
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Probable transcription factor.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- PTM: Phosphorylated by AKT1; insulin-induced (By similarity).
 CC -!- DISEASE: Involved in t(2;13)(q35;q14) and t(1;13)(p36;q14)
 CC chromosomal translocations in alveolar rhabdomyosarcoma-2 (RMS2)
 CC that involves FOXO1A and PAX3 or PAX7. The resulting protein is a
 CC transcriptional activator.
 CC -!- SIMILARITY: Contains 1 fork-head domain.

RP PHOSPHORYLATION ON SER-315 BY SGK.
 RX MEDLINE=20584913; PubMed=11154281;
 RA Brunet A., Park J., Tran H., Hu L.S., Hemmings B.A., Greenberg M.E.;
 RT "Protein kinase SGK mediates survival signals by phosphorylating the
 RT forkhead transcription factor FKHL1 (FOXO3a).";
 RL Mol. Cell. Biol. 21:952-965(2001).
 CC -!- FUNCTION: Probable transcription factor that may trigger apoptosis
 CC by inducing the expression of genes that are critical for cell
 CC death.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- PTM: Phosphorylated by AKT1 and by SGK/SGK1.
 CC -!- DISEASE: Involved in secondary acute leukemias by a chromosomal
 CC translocation t(6;11)(q21;q23) that involves FOXO3A and MLL/MLL.
 CC -!- SIMILARITY: Contains 1 fork-head domain.
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/AF6q21D125.html".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF032886; AAC39592.1; -;
 CC EMBL; AL365509; CAC26821.1; ALT_SEQ.
 CC EMBL; BC020227; AAH20227.1; -;
 CC EMBL; BC021224; AAH21224.1; -;
 CC EMBL; AJ001589; CAA04860.1; -;
 CC EMBL; AJ001590; CAA04861.1; -;
 CC HSSP; O63245; 2HEF.
 CC TRANSFAC; T02938; -;
 CC Genew; HGNC:3821; FOXO3A.
 CC MIN; 602681; -;
 CC GO; GO:0005737; C:cytoplasm; TAS.
 CC GO; GO:0005634; C:nucleus; TAS.
 CC GO; GO:0006917; P:induction of apoptosis; TAS.
 CC GO; GO:0006366; P:transcription from Pol II promoter; TAS.
 CC InterPro; IPR001766; TF_Fork_head.
 CC Pfam; PF00250; Fork_head; 1.
 CC PRINTS; PR00053; FORKHEAD.
 CC ProDom; PD000425; TF_Fork_head; 1.
 CC SMART; SM00339; FH; 1.
 CC PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
 CC PROSITE; PS00658; FORK_HEAD_2; 1.
 CC PROSITE; PS00339; FORK_HEAD_3; 1.
 CC Transcription regulation; DNA-binding; Nuclear protein; Apoptosis;
 CC Chromosomal translocation; Proto-oncogene; Phosphorylation.
 CC FT DNA_BIND 157 251
 CC FT MOD_RES 253 253
 CC FT MOD_RES 253 253
 CC FT MOD_RES 315 315
 CC FT CONFLICT 156 163
 CC FT CONFLICT 238 246
 CC FT CONFLICT 253 253
 CC FT CONFLICT 271 271
 CC FT CONFLICT 292 330
 CC FT CONFLICT 345 361
 CC FT CONFLICT 367 371
 CC FT CONFLICT 371 371
 CC FT CONFLICT 382 383
 CC FT CONFLICT 673 AA; E5B4E830665A9982 CRC64;
 CC SEQUENCE 673 AA; 71276 MW; 55B4E830665A9982 CRC64;
 CC
 CC Query Match 79.4%; Score 440.5; DB 1; Length 673;
 CC Best Local Similarity 77.5%; Pred. No. 3.3e-41;

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC      ENBL; U02310; AAA03629.1; -.
CC      ENBL; AF012885; AAC39591.1; -.
CC      ENBL; BC021981; AAH21981.1; -.
CC      PR; S40521; 2HFH.
CC      HSSP; Q63245; 2HFH.
CC      TRANSFAC; T02936; -.
CC      Genew; HGNC:3819; FOXO1A.
CC      MIM; 136533; -.
CC      GO; GO:0003700; F:transcription factor activity; TAS.
CC      GO; GO:0006916; P:anti-apoptosis; TAS.
CC      GO; GO:0006357; P:regulation of transcription from Pol II pro. . . ; TAS.
CC      InterPro; IPR001766; TF_Fork_head.
CC      Pfam; PF00250; Fork head; 1.
CC      PRINTS; PR00053; FORKHEAD.
CC      ProDom; PD000425; TF_Fork_head; 1.
CC      SMART; SM00339; FH; 1.
CC      PROSITE; PS00657; FORK HEAD 1; FALSE_NEG.
CC      PROSITE; PS00658; FORK HEAD 2; 1.
CC      PROSITE; PS00339; FORK HEAD 3; 1.
CC      Transcription regulation; DNA-binding; Nuclear protein;
CC      Phosphorylation; Chromosomal translocation; Proto-oncogene.
CC      DOMAIN 91 102 POLY-ALA.
CC      DOMAIN 120 130 POLY-PRO.
CC      DOMAIN 152 155 POLY-SER.
CC      DNA_BIND 159 235 FORK-HEAD.
CC      MOD_RES 256 256 PHOSPHORYLATION (BY PKB/AKT1) (BY
CC      SIMILARITY).
CC      CONFLICT 131 131 V -> L (IN REF. 2).
CC      SEQUENCE 655 AA; 69647 MW; 6DF6C994E740399 CRC64;

Query Match
Best Local Similarity 77.0%; Score 427.5; DB 1; Length 655;
Matches 75; Conservative 18; Mismatches 8; Indels 1; Gaps 1;

QY 1 KKTTRNAGNMSYAEILTTAIMASPEKRLTLAQVYEMVQNPYFRDKGDSNSAGWK 60
Db 151 KSSRRNAGNLSYADLITKAIESAEKRLTSLQIYEMVQNPYFRDKGDSNSAGWK 210

QY 61 NSIRNLSLHSFMRQNEGAGKSSWWVNPDP-AKPGMPPR 101
Db 211 NSIRNLSLHSFIRVQNEGTGKSSWWMLNPEGGKSGKSPRR 252

RESULT 3
FX01_MOUSE
ID FX01_MOUSE STANDARD; PRT; 652 AA.
AC Q9RIE0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Forkhead box protein O1A (Forkhead in rhabdomyosarcoma).
GN FOXO1A OR FOXO1 OR FKHR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99278356; PubMed=10347145;
RA Nakae J., Park B.C., Accili D.;
RT "Insulin stimulates phosphorylation of the forkhead transcription
RT factor FKHR on serine 253 through a Wortmannin-sensitive pathway.";
RL J. Biol. Chem. 274:15982-15985(1999).
CC -!- FUNCTION: Probable transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

```

```

CC      -!- PTM: PHOSPHORYLATED BY AKT; INSULIN-INDUCED.
CC      -!- SIMILARITY: Contains 1 fork-head domain.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).


EMBL; AF126056; AAD40636.1; -.
HSSP; Q63245; 2HFH.
MGD; MGI:1890077; Foxo1.
InterPro; IPR001766; TF_Fork_head.
Pfam; PF00250; Fork head; 1.
PRINTS; PR00053; FORKHEAD.
ProDom; PD000425; TF_Fork_head; 1.
SMART; SM00339; FH; 1.
PROSITE; PS00657; FORK HEAD 1; FALSE_NEG.
PROSITE; PS00658; FORK HEAD 2; 1.
PROSITE; PS00339; FORK HEAD 3; 1.
Transcription regulation; DNA-binding; Nuclear protein;
Phosphorylation.
DOMAIN 89 96 POLY-ALA.
DOMAIN 135 139 POLY-ALA.
DNA_BIND 156 232 FORK-HEAD.
MOD_RES 253 253 PHOSPHORYLATION (BY PKB/AKT1).
SEQUENCE 652 AA; 69502 MW; 3FE4C322AA85205F CRC64;

Query Match
Best Local Similarity 76.8%; Score 426.5; DB 1; Length 652;
Matches 75; Conservative 18; Mismatches 8; Indels 1; Gaps 1;

QY 1 KKTTRNAGNMSYAEILTTAIMASPEKRLTLAQVYEMVQNPYFRDKGDSNSAGWK 60
Db 148 KTSRRNAGNLSYADLITKAIESAEKRLTSLQIYEMVQNPYFRDKGDSNSAGWK 207

QY 61 NSIRNLSLHSFMRQNEGAGKSSWWVNPDP-AKPGMPPR 101
Db 208 NSIRNLSLHSFIRVQNEGTGKSSWWMLNPEGGKSGKSPRR 249

RESULT 4
FX04_MOUSE
ID FX04_MOUSE STANDARD; PRT; 505 AA.
AC Q9WH3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative fork head domain transcription factor AFX1 (Afxh) (Forkhead
DE box protein O4).
GN MLLT7 OR AFX1 OR AFX OR FOXO4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Biggs W.H. III, Cavenee W.K., Arden K.C.;
RT "Identification and characterization of murine members of the FKHR
RT subclass of winged-helix transcription factors.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Furuyama T., Nakazawa T., Mori N.;
RT "Mouse AFX, a forkhead type transcription factor.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Plays a role in the insulin signaling pathway (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: Contains 1 fork-head domain.


```

FT	23	CONFLICT	54	CONFLICT
FT	74	CONFLICT	74	P → S (IN REF. 2).
FT	79	CONFLICT	79	A → G (IN REF. 2).

```

FT CONFLICT 109 109 L -> F (IN REF. 2).
SQ CONFLICT 422 422 R -> P (IN REF. 1).
SQ SEQUENCE 505 AA; 53757 MW; 809CED90E6EFCAB8 CRC64;

Query Match
Best Local Similarity 74.0%; Score 410.5; DB 1; Length 505;
Matches 73; Conservative 15; Mismatches 13; Indels 1; Gaps 1;

QY 1 KKTTRNANWNNYSYAEILITTAIMASPEKRLTIAQVYEWVQVYPRDKGDSNSSAGWK 60
Db 92 RKGSRRNANGNSYAEILITTAIMASPEKRLTIAQVYEWVQVYPRDKGDSNSSAGWK 151

QY 61 NSTRHNLSLHSRFRMRQNEGAGKSSWVWVNPDKAKPGMNPTR 101
Db 152 NSIRHNLSLHSKFTIKVNEATGKSSWVWVNPDKAKPGMNPTR 193

RESULT 6
YA41_HUMAN
ID_YA41_HUMAN STANDARD; PRT; 622 AA.
AC Q9UPW0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative forkhead box protein KIAA1041.
GN KIAA1041.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirotsawa M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205 (1999).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB028964; BAA82993.1; -.
CC DR HSSP; Q63245; 2HPH.
CC DR TRANSFAC; T04179; -.
CC DR InterPro; IPR001766; TF Fork head.
CC DR Pfam; PF00250; Fork head; 1.
CC DR PRINTS; PR00053; FORKHEAD.
CC DR ProDom; PD000425; TF Fork_head; 1.
CC DR SMART; SM00339; FH; 1.
CC DR PROSITE; PS00657; FORK HEAD 1; 1.
CC DR PROSITE; PS00658; FORK HEAD 2; 1.
CC DR PROSITE; PS00339; FORK HEAD 3; 1.
CC KW Hypothetical protein; Transcription regulation; DNA-binding;
CC KW Nuclear protein.
CC FT DNA BIND 77 168 FORK-HEAD.
CC SEQUENCE 622 AA; 68928 MW; F8C57D71A7980FF CRC64;

Query Match
Best Local Similarity 40.1%; Score 222.5; DB 1; Length 622;
Matches 45; Conservative 19; Mismatches 21; Indels 11; Gaps 4;

QY 14 SYAELITTAIMASPEKRLTIAQVYEWVQVYPRDKGDSNSSAGWKSIRHNLHLHSRF 73
Db 152 SYAELITTAIMASPEKRLTIAQVYEWVQVYPRDKGDSNSSAGWKSIRHNLHLHSRF 172

RESULT 7
FXJ2_MOUSE
ID_FXJ2_MOUSE STANDARD; PRT; 565 AA.
AC Q9ES18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Forkhead box protein J2 (Fork head homologous X).
GN FOXJ2 OR FXH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20480369; PubMed=11025217;
RA Granadino B., Arias-de-la-Fuente C., Perez-Sanchez C., Parraga M.,
RA Lopez-Fernandez L.A., del Mazo J., Rey-Campos J.;
RT "Fhx (Foxj2) expression is activated during spermatogenesis and very
RT early in embryonic development.";
RL Mech. Dev. 97:157-160 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Salivary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Altshuler R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Katschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ussid T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson K.D., Mullahy S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield V.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Transcriptional activator. Able to bind to two different
CC type of DNA binding sites.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF253052; AAC30406.1; -.
CC DR EMBL; BC040395; AAH40395.1; -.
CC DR HSSP; Q63245; 2HPH.
CC DR TRANSFAC; T04170; -.
CC DR MGD; MGI:1926805; Foxj2.
CC DR InterPro; IPR001766; TF Fork head.
CC DR Pfam; PF00250; Fork_head; 1.
CC DR PRINTS; PR00053; FORKHEAD.

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF155132; AAF65927.1; -
DR EMBL; AF155133; AAK49016.1; -
DR EMBL; AL161978; CAB82315.1; -
DR PIR; T47161; T47161.
DR HSSP; Q63245; 2HFH.
DR TRANSFAC; T04169; -
DR TRANSFAC; T04171; -
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00659; FORK_HEAD_3; 1.
DR Transcription regulation; DNA-binding; Nuclear protein; Activator;
KW Alternative splicing
KW DNA_BIND 66 143 FORK-HEAD.
FT DOMAIN 266 270 POLY-SER.
FT DOMAIN 291 294 POLY-GLN.
FT DOMAIN 295 298 POLY-PRO.
FT DOMAIN 299 306 POLY-GLN.
FT DOMAIN 313 321 POLY-GLN.
FT DOMAIN 390 395 POLY-PRO.
FT DOMAIN 513 574 VNSYGHFQAPHLFPGSPMYPIPTQDSAGYNRPAAHVMVPRP
FT VARSPLIC 513 574 SVPPGANEIIPDDFDWDLIT -> GTAPQLPWRWLC
FT (in isoform FOXJ2.S).
FT /FTID=VSP 001544.
FT
FT
FT
SQ SEQUENCE 574 AA; 62394 MW; 258120EAE4B11EB CRC64;

Query Match 36.3%; Score 201.5; DB 1; Length 574;
Best Local Similarity 44.4%; Pred. No. 1.1e-14;
Matches 40; Conservative 18; Mismatches 23; Indels 9; Gaps 3;

Qy 14 SYAELITTAIMASPEKELTIAQVYVWVQVYPRDKGDSNSAGWKNISIRHNLHLHSRF 73
Db 70 STATLITYAINSPAKMTLSEIYRNICDNFFPYK-----NAGIGWKNISIRHNLHLNKCF 124

Qy 74 MRI--QNEGAGKSSWWVINPDAPGNMRR 101
Db 125 RKVPRDRDPGKGSYWTI--DTCEDISKRR 152

RESULT 9
FXJ2 HUMAN STANDARD; PRT; 351 AA.
ID -FXJ2 HUMAN STANDARD; PRT; 351 AA.
AC Q12951; O14518; Q8N6L8;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Forkhead box protein 11 (Forkhead-related protein FKHL10) (Forkhead-
DE related transcription factor 6) (FREAC-6) (Hepatocyte nuclear factor 3
DE forkhead homolog 3) (HNF-3/fork-head homolog-3) (HFH-3).
GN FOXJ1 OR FKHL10 OR FREAC6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN [1] SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=Kidney;
RC MEDLINE=97298077; PubMed=9153225;
RA Overdier D.G., Ye H., Peterson R.S., Clevidence D.E., Costa R.H.;
RT "The winged helix transcriptional activator HFH-3 is expressed in the
RT distal tubules of embryonic and adult mouse kidney."
RT J. Biol. Chem. 272:13725-13730(1997).
RN [2]

DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00659; FORK_HEAD_3; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Activator.
FT DNA_BIND 66 143 FORK-HEAD.
FT DOMAIN 266 270 POLY-SER.
FT DOMAIN 291 294 POLY-GLN.
FT DOMAIN 295 298 POLY-GLN.
FT DOMAIN 306 314 POLY-GLN.
SQ SEQUENCE 565 AA; 61569 MW; 9178AFF3F9227AD4 CRC64;

Query Match 36.3%; Score 201.5; DB 1; Length 565;
Best Local Similarity 44.4%; Pred. No. 1.1e-14;
Matches 40; Conservative 18; Mismatches 23; Indels 9; Gaps 3;

Qy 14 SYAELITTAIMASPEKELTIAQVYVWVQVYPRDKGDSNSAGWKNISIRHNLHLHSRF 73
Db 70 STATLITYAINSPAKMTLSEIYRNICDNFFPYK-----NAGIGWKNISIRHNLHLNKCF 124

Qy 74 MRI--QNEGAGKSSWWVINPDAPGNMRR 101
Db 125 RKVPRDRDPGKGSYWTI--DTCEDISKRR 152

RESULT 8
FXJ2 HUMAN STANDARD; PRT; 574 AA.
ID -FXJ2 HUMAN STANDARD; PRT; 574 AA.
AC Q9P0K8; Q96PS9; Q9NSN5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 18-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Forkhead box protein J2 (Fork head homologous X).
GN FOXJ2 OR FHX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN [1] SEQUENCE FROM N.A. (ISOFORM FOXJ2.L).
RP MEDLINE=20239944; PubMed=10777590;
RA Perez-Sanchez C., Gomez-Ferreria M.A., de la Fuente C.A.,
RA Granadino B., Velasco G., Esteban A., Rey-Campos J.;
RT "FHX, a novel fork head factor with a dual DNA binding specificity."
RT J. Biol. Chem. 275:12909-12916 (2000).
RN [2]
RN [3] SEQUENCE FROM N.A. (ISOFORM FOXJ2.S).
RP MEDLINE=20425082; PubMed=10966786;
RA Perez-Sanchez C., de la Fuente C.A., Gomez-Ferreria M.A.,
RA Granadino B., Rey-Campos J.;
RT "FHX.L and FHX.S, two isoforms of the human fork-head factor FHX
RT (FOXJ2) with differential activity."
RT J. Mol. Biol. 301:795-806(2000).
RN [3]
RP SEQUENCE OF 458-574 FROM N.A. (ISOFORM FOXJ2.L).
RC TISSUE=Melanoma;
RA Anorge W., Winkner U., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transcriptional activator. Able to bind to two different
CC type of DNA binding sites. Isoform FOXJ2.L behaves as a more
CC potent transactivator than FOXJ2.S.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=FOXJ2.L; Synonyms=FHX.L;
CC IsoId=Q9P0K8-1; Sequence=Displayed;
CC Name=FOXJ2.S; Synonyms=FHX.S;
CC IsoId=Q9P0K8-2; Sequence=VSP 001544;
CC -!- TISSUE SPECIFICITY: Widely expressed.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

[illegible]


```

RX MEDLINE=92265309; PubMed=1350202;
RA Tao W., Lai E.;
RT "Telencephalon-restricted expression of BF-1, a new member of the
RT HNF-3/fork head gene family, in the developing rat brain.";
RL Neuron 8:957-966(1992).
CC -!- FUNCTION: Plays an important role in the establishment of the
CC regional subdivision of the telencephalon. Sequence-specific DNA-binding
CC protein with a distinct binding specificity.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Is expressed in the cortex, olfactory bulb,
CC hippocampus, and caudate putamen.
CC -!- DEVELOPMENTAL STAGE: More abundant in fetal brain compared with
CC the adult brain.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M87634; AAA40812.1; -.
CC PIR; JH0672; JH0672.
CC HSSP; Q63245; 2HFH.
CC TRANSFAC; T01831; -.
CC InterPro; IPR001766; TF_Fork_head.
CC PRINTS; PR00053; FORKHEAD.
CC ProDom; PD000425; TF_Fork_head; 1.
CC SMART; SM00339; FH; 1.
CC PROSITE; PS00657; FORK_HEAD_1; 1.
CC PROSITE; PS00658; FORK_HEAD_2; 1.
CC PROSITE; PS00039; FORK_HEAD_3; 1.
CC Transcription regulation; DNA-binding; Nuclear protein;
CC Developmental protein.
CC DOMAIN 33 56 HIS-RICH.
CC DOMAIN 57 82 PRO-RICH.
CC DOMAIN 70 74 POLY-GLN.
CC DNA_BIND 171 262 FORK-HEAD.
CC SEQUENCE 480 AA; 51466 MW; 00425B2F5298F444 CRC64;
CC
CC Query Match 34.0%; Score 188.5; DB 1; Length 480;
CC Best Local Similarity 43.9%; Pred. No. 2.5e-13;
CC Matches 36; Conservative 20; Mismatches 19; Indels 7; Gaps 2;
CC
CC QY 14 SYAELITTAIASPEKRLTIAQVYEMVYQVYPRDKGDSNSAGWKNISRHNLSLHSRF 73
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC DB 176 SYNALIMMAIRQSPEKRLTNGIYEFIMKNPFYRE-----NKQGWQNSIRHNLKCF 230
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC QY 74 MRIGN--EGAGKSGWVINPDA 93
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC DB 231 VKVPRHYDDPGKGYWMLDPSS 252
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC
CC RESULT 15
CC FXGB_MOUSE
CC ID FXGB_MOUSE STANDARD; PRT; 481 AA.
CC AC Q60987;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE Forkhead box protein G1B (Forkhead-related protein FKHL1)
CC DE (Transcription factor BF-1) (Brain factor 1) (BF1).
CC GN FOXGLB OR FOXG1 OR FKHL1 OR HFHBF1.
CC OS Mus musculus (Mouse)
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC OX NCBI_TaxID=10090;
CC RN [1]
CC RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=96346618; PubMed=8738140;
RA Li H., Tao W., Lai E.;
RT "Characterization of the structure and function of the gene for
RT transcription factor BF-1, an essential regulator of forebrain
RT development.";
RL Brain Res. Mol. Brain Res. 37:96-104(1996).
RN [2]
RN SEQUENCE FROM N.A.
RP
RX TISSUE:Olfactory epithelium;
RX MEDLINE=23388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins B., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max A.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Plays an important role in the establishment of the
CC regional subdivision of the developing brain and in the
CC development of the telencephalon. Sequence-specific DNA-binding
CC protein with a distinct binding specificity (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: CNS, and nasal half of the retina.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U36760; AAB42158.1; -.
CC EMBL; BC046958; AAB46958.2; -.
CC HSSP; Q63245; 2HFH.
CC TRANSFAC; T02353; -.
CC MGD; MGI:1347464; Foxg1.
CC GO; GO:0007420; P:brain development; IMP.
CC GO; GO:0030182; P:neuron differentiation; IMP.
CC GO; GO:0007346; P:regulation of mitotic cell cycle; IMP.
CC InterPro; IPR001766; TF_Fork_head.
CC Pfam; PF00250; Fork_head; 1.
CC PRINTS; PR00053; FORKHEAD.
CC ProDom; PD000425; TF_Fork_head; 1.
CC SMART; SM00339; FH; 1.
CC PROSITE; PS00657; FORK_HEAD_1; 1.
CC PROSITE; PS00658; FORK_HEAD_2; 1.
CC PROSITE; PS00039; FORK_HEAD_3; 1.
CC Transcription regulation; DNA-binding; Nuclear protein;
CC Developmental protein.
CC DOMAIN 33 56 HIS-RICH.
CC DOMAIN 57 83 PRO-RICH.
CC DOMAIN 70 75 POLY-GLN.
CC DNA_BIND 172 263 FORK-HEAD.
CC SEQUENCE 481 AA; 51624 MW; FF6786C3513AB452 CRC64;
CC
CC Query Match 34.0%; Score 188.5; DB 1; Length 481;
CC Best Local Similarity 43.9%; Pred. No. 2.5e-13;
CC Matches 36; Conservative 20; Mismatches 19; Indels 7; Gaps 2;

```

Qy 14 SYAELITITAIMASPEKRLTLAQVYEWVQNVYPERDKGDSNSSAGWKNSIRHNLSLHSRF 73
Db 177 SYNALIMWAIHQSEPEKRLTLNGIYEFIMKNFFYRE-----NKOGWQNSIRHNLSLNKCF 231
Qy 74 MRION--EGAGKSSWWVINPDA 93
Db 232 VKVPRHYDDPGKGNWMLDPSS 253

Search completed: July 23, 2004, 10:24:37
Job time : 14 secs

This Page Blank (uspto)